University of Alberta

Speech Intelligibility Probe for Children with Cleft Palate Version 3: Assessment of Reliability and Validity

by

Carrie Lynne Gotzke ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Science

in

Speech Language Pathology

Department of Speech Pathology and Audiology

Edmonton, Alberta

Fall 2005

Library and
Archives Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

0-494-09176-2

Your file Votre référence
ISBN:
Our file Notre reterence
ISBN:

NOTICE:
The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:
L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

ABSTRACT

The purpose of this study was to evaluate the reliability and validity of the Speech Intelligibility Probe for Children with Cleft Palate (SIP-CCLP) Ver. 3. Participants, aged 39 to 84 months, included 12 children with and 24 children without cleft palate. SIP-CCLP imitative word productions and a 100-word spontaneous speech sample were recorded and played back to listeners who completed open-set and closed-set response tasks. The children with cleft palate had significantly lower intelligibility scores, fewer phonetic contrast items correct, greater percent correct/distorted items and fewer obstruents correct than age-similar children without cleft palate. A developmental progression in SIP-CCLP scores was also found. Intelligibility scores from the SIP-CCLP and the spontaneous sample were positively correlated ($r=.79$, $p<.000$). PCC was similar across sampling conditions. Reliability was acceptable for all measures. The results support the validity and reliability of the SIP-CCLP as a clinical outcome measure for children with cleft palate.

Acknowledgements

This work was funded by the National Science and Engineering Research Council and the Canadian Language and Literacy Research Network.

I would like to express my gratitude to my supervisor, Dr. Megan Hodge. I cannot begin to define all that I have learned from you or to express how much I appreciated the time, effort and encouragement you offered during this process. My thesis has also been made stronger through the guidance of my committee members: Karen Pollock, Terry Nearey, Sharon Hundert and Sheila Ennis. Thank you for your support. I would also like to thank Ralph Shelton for his comments.

I would also like to express my thanks to Kimberley Cote and Kari Dul for lab assistance and for always finding me "just one more listener." As well, thanks to Jason Daniels for software development.

This thesis would not have been possible without the support of the Alberta Children's Hospital, Stollery Children's Hospital and Glenrose Rehabilitation Hospital. Special thanks to Cindy Humphrey for her commitment to recruitment and for making room for me on busy clinic days. I would also like to extend my thanks to the children and their families and all the listeners who participated in this project.

This thesis is dedicated to my parents, Ryan and Marilyn Gotzke.

Table of Contents

INTRODUCTION **1**

RATIONALE 1
LITERATURE REVIEW 2
Speech Intelligibility *2*
Speech of Children with Cleft Palate *5*
Nature of Cleft Palate 5
Velopharyngeal Inadequacy Related to Speech 8
Relationship between Speech and Expressive Language 12
Speech Characteristics of Children with Syndromic Cleft Palate 13
Speech Intelligibility 15
Speech Intelligibility Probe for Children with Cleft Palate Version 2 .... *18*
Development of SIP-CCLP Ver. 2 18
Description of SIP-CCLP Ver. 2 Contrast Pair Development 22
Cleft-related errors 23
Developmental errors 24
Unknown errors 24
Description of SIP-CCLP Ver. 2 Administration 25
Validity and Reliability of SIP-CCLP Ver. 2 27
Content validity. 28
Construct validity. 29
Criterion validity. 34
Reliability. 36
Recommended Changes to Revise SIP-CCLP Ver. 2 to Version 3 36
SIP-CCLP Ver.3 closed-set error coding. 37
SIP-CCLP Ver. 3 closed-set analysis software 38
SIP-CCLP Ver. 3 closed-set error profiling and graphing. 39
SIP-CCLP Ver. 3 closed-set response task reliability 39
Summary. 40
PURPOSE 40
Research Questions and Hypotheses *40*
Reliability 40
Construct Validity 41
Criterion Validity 42

METHOD **44**

RESEARCH DESIGN 44
CHILD PARTICIPANTS 44
Recruitment *44*
Inclusion Criteria *46*
Children with Cleft Palate 46
Children without Cleft Palate 49
SPEECH SAMPLING CONDITIONS 50
Spontaneous Sample *50*
SIP-CCLP Ver. 3 .... *51*

ASSESSMENT OF VALIDITY AND RELIABILITY 52
Preparation of Recordings for Listener Judges *52*
Spontaneous Sample 52
SIP-CCLP Ver. 3 54
Listener Judges *54*
Recruitment 54
Instructions 55
SIP-CCLP Ver. 3 word identification – open-set response task 56
SIP-CCLP Ver. 3 phonetic contrast identification – closed-set response task 57
Spontaneous sample 58
Measurement of Dependent Variables *58*
Spontaneous Sample 59
Intelligibility analysis 59
Phonetic analysis 59
SIP-CCLP Ver. 3 60
Intelligibility analysis 60
Phonetic analysis 62
Closed-set: Phonetic contrast error analysis 63
Closed-set: Sound analysis of error patterns 66
Closed-set: Distortion analysis 67
Analysis *68*
Reliability 72
SIP-CCLP open-set response task 72
SIP-CCLP closed-set response task 72
Construct Validity 72
Children with and without cleft palate 72
Children without cleft palate of different ages 73
Criterion Validity 73

RESULTS 74

RELIABILITY AND CONSTRUCT VALIDITY 74
Comparison of Children with and without Cleft Palate *74*
Intelligibility Analysis 74
SIP-CCLP open-set response task 74
SIP-CCLP closed-set response task 75
Error Analysis: Closed-set Response Task 77
Sound Analysis of Errors: Closed-set Response Task 83
Sound analysis of errors organized by manner 84
Sound analysis of errors organized by place of articulation 86
Sound analysis of errors organized by voicing 87
Distortion Analysis: Closed-set Response Task 88
Comparison of Children without Cleft Palate by Age Group *90*
Intelligibility Analysis 90
SIP-CCLP open-set response task 90
SIP-CCLP closed-set response task 92
Error Analysis: Closed-set Response Task 93
CRITERION VALIDITY 98

Intelligibility ... *98*
Comparison of Children with and without Cleft Palate ... 98
Comparison of Children without Cleft Palate by Age Group ... 102
Phonetic Analysis ... *103*
Children with and without Cleft Palate ... 103

DISCUSSION ... 106

RELIABILITY ... 106
Identification Tasks ... *106*
Ratings of Distorted/Clear ... *108*
Conclusions ... *109*
CONSTRUCT VALIDITY ... 109
Comparison of Children with and without Cleft Palate ... *109*
SIP-CCLP Open-set Response Task ... 109
SIP-CCLP Closed-set Response Task ... 110
Intelligibility score and percent phonetic contrast items correct ... 110
Error pattern analysis ... 111
Sound analysis of error patterns ... 117
Distortion analysis ... 118
Conclusions ... 124
Comparison of Children without Cleft Palate by Age Group ... *125*
CRITERION VALIDITY ... 126
Intelligibility ... *126*
Phonetic Analysis ... *128*
RECOMMENDATIONS FOR FUTURE REVISIONS TO SIP-CCLP ... 130
Number of Target Phonetic Contrast Items ... *130*
Analysis and Graphing Software ... *135*
Summary of Recommended Revisions For SIP-CCLP Ver.4 ... *135*
RESULTS FOR TWO CHILDREN WITH SYNDROMIC CLEFT PALATE ... 135
FUTURE RESEARCH ... 142
CONCLUSIONS ... 143

REFERENCES ... 145

APPENDICES

A. SIP-CCLP VER. 2 STIMULUS WORDS ... 152
B. SIP-CCLP VER. 3 TARGET PHONETIC CONTRAST PAIRS (CLOSED-SET RESPONSE TASK) ... 153
C. PRESENTATION SCREEN FOR SIP-CCLP VER. 2 STIMULUS WORDS ... 160
D. PRESENTATION SCREEN FOR SIP-CCLP VER. 2 OPEN-SET RESPONSE TASK ... 161
E. PRESENTATION SCREEN FOR SIP-CCLP VER. 2 CLOSED-SET RESPONSE TASK INSTRUCTIONS ... 162
F. PRESENTATION SCREEN FOR SIP-CCLP VER. 2 CLOSED-SET RESPONSE TASK ... 163
G. POWER ANALYSIS FOR ESTIMATING NUMBER OF SUBJECTS REQUIRED ... 164
H. INFORMATION LETTER AND CONSENT FORM FOR CHILDREN ... 166
I. INFORMATION LETTER AND CONSENT FORM FOR LISTENER JUDGES ... 169
J. PRIMING WORD LIST FOR SIP-CCLP VER. 3 ... 172

K. LIST OF DEPENDENT VARIABLES: CONSTRUCT AND CRITERION VALIDITY 173
L. LIST OF DEPENDENT VARIABLES: RELIABILITY 174
M. UNCLASSIFIED ERRORS FOR CHILDREN WITH CLEFT PALATE FROM SIP-CCLP CLOSED-SET RESPONSE TASK 176
N. ERRORS IDENTIFIED IN THE SIP-CCLP CLOSED-SET TASK FOR THE CHILDREN WITH AND WITHOUT CLEFT PALATE 177
O. SIP-CCLP CLOSED-SET RESPONSE TASK ERRORS FOR THE CHILDREN WITHOUT CLEFT PALATE, ORGANIZED BY AGE GROUP 181
P. COMPARISON OF ERRORS PATTERNS TESTED IN CURRENT STUDY AND WHITEHILL AND CHAU (2004 183

List of Tables

Table 1.
Development of SIP-CCLP Ver. 2 Administration: Child Recordings 21

Table 2.
Development of SIP-CCLP Ver. 2 Administration: Listener Response Tasks 22

Table 3.
Child Participant Characteristics for Gotzke (2003) 29

Table 4.
SIP-CCLP and Spontaneous Speech Sample Intelligibility and Closed-set Analysis Error Scores for Children with and without Cleft Palate 32

Table 5.
Subject Pool Characteristics for Children Recruited for Feltz et al. (2002), Gotzke (2003) and the Current Study 48

Table 6.
Subject Information for Children with Cleft Palate and their Age-Similar Peers without Cleft Palate 70

Table 7.
Subject Information for Children without Cleft Palate 71

Table 8.
Frequency of Cleft-Related, Developmental and Unknown Errors by SIP-CCLP Error Subcategories 82

Table 9.
SIP-CCLP and Spontaneous Sample Scores for the Four Age-Groups of Children without Cleft Palate 91

Table 10.
Frequency of Cleft-Related, Developmental and Unknown Errors by SIP-CCLP Error Subcategories 98

Table 11.
Suggested New Target Phonetic Contrast Items for the Next Version of SIP-CCLP (Ver. 4) 132

Table 12.
Frequency Distribution of the Phonetic Content of SIP-CCLP Stimulus Words, the Targets in the Closed-Set Contrast Items for the Durrent and Proposed Versions and of Conversational Speech for Children in Grade 1 134

Table 13.
SIP-CCLP Ver. 3 and Spontaneous Sample Scores for Two Children with Syndromic Cleft Palate and 12 Children with Nonsyndromic Cleft Palate 137

List of Figures

Figure 1.
Example of error pattern coding for SIP-CCLP Ver. 3 for the phonetic contrast pair "tail" (target) and "pail" (foil)........38

Figure 2.
Flowchart for selection of 100-word spontaneous sample........53

Figure 3.
Example of results for SIP-CCLP Ver. 3 closed-set response task........61

Figure 4.
Example of collated listener responses on the SIP-CCLP Ver. 3 closed-set response task........64

Figure 5.
Intelligibility scores for children with and without cleft palate for the SIP-CCLP open-set, SIP-CCLP closed-set and spontaneous sample open-set response tasks........77

Figure 6.
Percent phonetic contrast items correct by group........78

Figure 7.
Frequency of errors by listener response for each group of children........79

Figure 8.
Frequency of errors by category and group........81

Figure 9.
Contrast target error profile based on results from SIP-CCLP for the children with and without cleft palate........83

Figure 10.
Percent target obstruents and sonorants correct in SIP-CCLP closed-set response task for children with and without cleft palate........84

Figure 11.
Percent target sounds correct organized by manner in SIP-CCLP closed-set response task for children with and without cleft palate.....85

Figure 12.
Percent target sounds correct organized by place of articulation in SIP-CCLP closed-set response task for children with and without cleft palate.....86

Figure 13.
Percent target sounds correct organized by voicing in SIP-CCLP closed-set response task for children with and without cleft palate.....87

Figure 14.
Percent correct distorted and incorrect distorted items on SIP-CCLP closed-set response task for children with and without cleft palate.....89

Figure 15.
Intelligibility scores on the SIP-CCLP open-set and closed-set response tasks, compared with the spontaneous sample intelligibility scores for the four age groups of children without cleft palate.....93

Figure 16.
Percent phonetic contrast items correct for the four age groups of children without cleft palate.....94

Figure 17.
Frequency of errors by listener response for the four age groups of children without cleft palate.....95

Figure 18.
Frequency of errors by category (cleft-related, developmental, unknown and unclassified) and by age group for the children without cleft palate.....97

Figure 19.
Relationship between SIP-CCLP open-set and spontaneous sample intelligibility scores for the children with and without cleft palate.....100

Figure 20.
Relationship between SIP-CCLP open-set and spontaneous sample intelligibility scores for the children with cleft palate.....101

Figure 21.
Relationship between SIP-CCLP open-set and spontaneous sample intelligibility scores for the children without cleft palate.....101

Figure 22.
Relationship between SIP-CCLP open-set and spontaneous sample intelligibility scores for the 24 children without cleft palate.....103

Figure 23.
Percent consonants correct for the children with and without cleft palate for SIP-CCLP closed set response task, the narrow transcription of the SIP-CCLP words and the narrow transcription of the 100-word spontaneous sample.....105

Figure 24.
Accuracy scores for age-similar children with and without cleft palate.....121

Figure 25.
Relationship between SIP-CCLP closed-set intelligibility scores and accuracy scores for children with and without cleft palate.....122

Figure 26.
Relationship between SIP-CCLP open-set intelligibility scores and accuracy scores for children with and without cleft palate.....123

Figure 27.
Relationship spontaneous sample intelligibility scores and accuracy scores for children with and without cleft palate.....124

Figure 28.
Relationship between SIP-CCLP open-set and spontaneous sample intelligibility for all children with cleft palate (including children with a diagnosed syndrome) divided into Group1 (< 12% difference between intelligibility scores) and Group 2 (>20% difference between intelligibility scores).....140

Figure 29.
Contrast target error profile based on results from SIP-CCLP Ver. 3 for CP04 and CP12 141

INTRODUCTION

Rationale

Assessment of intelligibility is often included as part of the evaluation procedure for children with cleft palate. Therefore, a reliable and valid means of assessing intelligibility for this population is essential. Furthermore, it is important that an intelligibility measure is sensitive to the particular error patterns of the population for which it is to be used (Kent, Weismer, Kent, & Rosenbek, 1989).

The Speech Intelligibility Probe for Children with Cleft Palate Version 1 (SIP-CCLP Ver. 1) is a measure of single word intelligibility that was developed based on the expected speech error patterns of children with cleft palate (Connolly, 2001; Feltz, McClure, & O'Hare, 2002). The presentation format of SIP-CCLP Ver. 1 was revised from a PowerPoint (Microsoft, 2000) presentation to a customized software program using Authorware 6.0 (Macromedia, 2001) and renamed SIP-CCLP Ver. 2.

Gotzke (2003) conducted a preliminary evaluation of the reliability and validity of SIP-CCLP Ver. 2 with 12 children who had language development within the normal range. Eight children had typical speech development and no history of speech mechanism impairment and four children had cleft palate. Based on the results from this small study, it was concluded that the SIP-CCLP Ver. 2 had potential to be a reliable and valid clinical tool. Revisions were also recommended for the next software version, i.e., SIP-CCLP Ver. 3. The purpose of the current study was to evaluate the validity and reliability of SIP-CCLP Ver. 3 using speech samples from a larger number of children, including those with and without cleft palate.

Literature Review

Speech Intelligibility

Intelligibility can be defined as the degree to which an individual's spoken message is recovered by a listener (Kent et al., 1989) and is considered to be a functional indicator of an individual's oral communication competence (Metz, Samar, Schiavetti, Sitler, & Whitehead, 1985). In clinical settings, speech intelligibility is used as a guide to determine if intervention is required, to set intervention goals and to evaluate the effectiveness of intervention (Gordon-Brannan & Hodson, 2000). Because of the importance of intelligibility in many clinical decisions, it is essential that intelligibility be measured using methods that are both accurate and valid (Gordon-Brannan & Hodson, 2000). However, there is a great deal of variability in clinical procedures used to determine a speaker's intelligibility.

Clinically, intelligibility is often based on an estimate of the percentage of words understood in a continuous spontaneous speech sample (Gordon-Brannan, 1994). From this percentage, the clinician makes subjective judgments about the intelligibility of a client. According to Gordon-Brannan (1994), this method is neither reliable nor valid.

Intelligibility may also be estimated through the use of rating scales. A literature review of intelligibility in cleft palate speakers conducted by Whitehill (2002) found that the majority of these studies used equal-appearing interval scales for assessing intelligibility. The validity of using interval scales for evaluating intelligibility has been questioned because listeners cannot divide intelligibility into equal intervals (Schiavetti, 1992; Whitehill, 2002). Furthermore, other speech variables, such as hypernasality, nasal emission or articulatory distortion, may make it difficult for listeners to focus only on intelligibility (Witzel, 1995).

Another approach to evaluating intelligibility relies on the ability of unfamiliar listeners to identify words spoken by an individual. The listener is required to orthographically transcribe a speech sample (i.e., open-set response task) and/or identify a word spoken from a word list (i.e., closed-set response task). The clinician then calculates a percentage for the number of words identified correctly. This value is described as the client's intelligibility score (Gordon-Brannan & Hodson, 2000). This method of measuring intelligibility is considered to be valid (Schiavetti, 1992).

An intelligibility score is not an absolute quantity. The test material, personnel, training and testing procedures, in addition to listener, speaker and environmental variables, all influence the intelligibility score obtained (Weston & Shriberg, 1992; Kent, Miolo, & Bloedel, 1994; Gordon-Brannan, 1994). The intelligibility score of a single speaker may vary depending on how these factors are manipulated and controlled. Consequently, intelligibility scores must be evaluated with respect to the manner in which they are collected (Kent et al., 1989).

Considerable attention has been given to the question of what effect sampling condition has on intelligibility scores. Generally, this debate has focused on the relative merits of using speech samples generated imitatively versus spontaneously and on using single word versus phrase or sentence stimuli when sampling speech. Controversy exists with respect to which method of sampling speech provides the most accurate representation of a speaker's abilities and which is the most appropriate for use in intelligibility assessment.

Studies comparing sampling conditions for intelligibility measures have reported varying results. Drozda, Gustafson, Roberts, Hodge, and Wellman (2002) found no significant difference in intelligibility scores between spontaneous and imitative sampling conditions, using phrase and sentence stimuli, for six children with and six children without motor speech disorders. However, Chin, Finnegan and Chung (2001) suggested that at finer levels of analysis, such as percent consonants correct, differences may be noted between spontaneous and imitative samples. Drozda et al. (2002) reported that scores for percent consonants correct were significantly higher in the spontaneous condition than the imitative condition for their groups of children with and without motor speech disorders. However, they did not find a significant difference between the sampling conditions for either group of children for percent vowels correct or percent syllable shape correct. This result is contrary to Johnson, Weston and Bain (2004) who found that percent consonants correct obtained from phonetic transcription of imitated sentences and a spontaneous sample were similar for children with speech delay.

Gordon-Brannan and Hodson (2000) found significant correlations among intelligibility scores for spontaneous speech, imitated words and imitated sentence conditions for prekindergarten children with phonologies varying from no speech sound errors to extensive omissions. A correlation of $r = 0.79$ ($p < .001$) was found between intelligibility scores for the spontaneous speech sample (open-set) and imitated words (closed-set), whereas a correlation of $r = 0.85$ ($p < .001$) was found between the spontaneous speech (open-set) and imitated sentence (open-set) conditions. However, the mean percentages for number of words identified correctly for the single-word imitation procedure were lower than those obtained for the spontaneous speech procedure (64%

and 76%, respectively). Mean percentages of number of words identified correctly differed only by 2% for the imitated sentences and spontaneous speech sample conditions. The authors stated that the word imitation procedure, whether single words or word combinations, failed to adequately reflect many of the factors that contribute to intelligibility (e.g., suprasegmental, contextual and linguistic cues). Stimley and Hambreacht (1999) found a weaker, nonsignificant positive correlation (r = .24, p > .05) between single word closed-set and conversational open-set speech intelligibility scores for 15 children with articulation or phonological disorders, suggesting that sampling condition does influence speech intelligibility measures. Both Gordon-Brannan and Hodson, and Stimley and Hambreacht recommended that an individual's overall intelligibility score be determined using spontaneous rather than imitative speech samples and open-set rather than closed-set judgment tasks.

Speech of Children with Cleft Palate

Nature of Cleft Palate

A cleft may be defined as an abnormal opening or a fissure in an anatomical structure that is normally closed (Kummer, 2001). Clefting occurs as a result of the failure of structures to fuse during embryological development (Kummer, 2001). Both cleft lip and cleft palate can occur. A cleft lip may be bilateral or unilateral. A cleft palate may extend through all or part of the hard and/or soft palate and depending on the length and location of the cleft, variably affects speech, swallowing and breathing. Many combinations of cleft lip and palate have been described. Both structure and function of the palate are affected when clefting has occurred.

A cleft of the hard palate may result in structural abnormalities of the alveolar ridge. The alveolar ridge is a place of articulation for many English consonant sounds including /t, d, n, s, z, l/. If the alveolar ridge is malformed, children may avoid using this place of articulation and instead use the lips, back of tongue, pharyngeal wall and/or larynx (Golding-Kushner, 1995). For example, children with cleft palate will often produce alveolars at a bilabial place of articulation such that "tail" becomes "pail." These sound substitutions or distortions will affect how well these children are understood. These phonetic errors may persist even after surgical repair of the cleft.

A fistula or hole may also be associated with a cleft of the hard palate. Fistulas allow air to pass into the nasal cavity, adversely affecting the ability of the child to build up sufficient air pressure in the oral cavity for production of obstruent consonants. Obstruent consonants or obstruents, which include stops, fricatives and affricates, are produced when airflow through the oral cavity is restricted via narrowing or complete occlusion. Air pressure is built up behind the site of constriction or closure. Forcing air through a narrow constriction or releasing the occlusion rapidly produces noise. Fricatives are produced when air is forced through the narrowed opening (i.e., /f, v, θ, ð, s, z, ʃ, ʒ ,h /), whereas stops are produced by rapid release of the closure (i.e., /p, b, t, d, k, g/). In production of affricates (i.e., /tʃ, ʤ/) complete occlusion of the oral cavity is followed by creation of a narrowed opening. If air escapes to the nasal cavity via the fistula, air pressure can not be built up in the oral cavity behind the site of restriction. As a result, obstruent consonants will sound weak or distorted. Air escape to the nasal cavity may be so great that stops sound like their nasal counterparts such that /b, d, g/ sound like /m, n, ŋ/. In both situations, the ability of listeners to identify obstruents correctly and therefore, the child's overall intelligibility will be affected negatively. To compensate for an inability to build up sufficient air pressure in the oral cavity during production of obstruents, a child with cleft palate may try to control the air stream at different places in the vocal tract where pressure build-up is possible. For example, the child may valve the air stream at the level of the larynx resulting in the production of "tape" as "ʔape." "Tape" would be produced with a glottal stop (ʔ) in initial position to differentiate it from "ape" with a vowel in initial position, thereby increasing the number of words that the child with cleft palate can use contrastively. Similar speech error patterns may result when the cleft extends into the soft palate. Once again, these error patterns may persist after the fistula or the cleft of the soft palate has been repaired surgically. The effect of a compromised velopharyngeal mechanism is discussed in greater detail in the following section.

A child with cleft palate may exhibit any or all of these speech error patterns depending on the location and extent of the cleft. When these errors persist after surgery to repair the cleft, these errors are considered to be phonologic. Phonologic errors occur when children adopt incorrect rules for organization and representation of the sound system (Chapman, 1993). In the case of cleft palate, these errors may develop originally in response to compromised articulatory structures and persist because the child continues to use the learned compensatory pattern as the rule for production even after the underlying anatomic deficit has been corrected (Chapman, 1993). As well, the speech of these children may contain developmental errors. Developmental errors are those that are seen in the speech of young children with typical speech development and no history of cleft lip and/or palate or other physical abnormality affecting speech behavior (Golding-Kushner, 1995; Gotzke, 2003).

Velopharyngeal Inadequacy Related to Speech

The velopharyngeal mechanism serves to separate the oral and nasal cavities during eating, drinking and speaking. It is composed of the velum or soft palate and its internal muscles, including the paired levator veli palatini muscles, the palatoglossus and musculus uvulae, as well as, muscles in the pharyngeal wall (e.g., superior pharyngeal constrictor). During production of obstruent sounds, the levator veli palatini muscles act to elevate the velum while the pharyngeal constrictors contract causing medial displacement of the pharyngeal walls. The combination of these two actions serves to form a seal between the oral and nasal cavities, preventing air flow through the nasal passage. For production of obstruents, tight closure of velopharyngeal mechanism is required to ensure adequate air pressure is built up behind the site of restriction. If tight closure is not attained, obstruent consonants may sound weak or distorted.

For another category of sounds, termed sonorants, tight velopharyngeal closure is not required. Sonorants are sounds produced with a relatively open vocal tract and include nasals (i.e., /m, n, ŋ/), liquids (i.e., /l, r/), glides (i.e., /w, j/) and vowels. During production of nasals, the velopharynx remains open and the oral cavity is occluded, such that air escapes out the nasal, not the oral, airway. During production of liquids, glides and vowels, the velopharynx is closed but not as tightly as with production of obstruents. As a result, some air may escape through the nasal cavity without reducing the clarity of the sound being produced.

The terms “velopharyngeal inadequacy”, “velopharyngeal insufficiency” and “velopharyngeal incompetence” have all been used in the literature to refer to instances in which the velopharyngeal valve is not functioning normally. To better define how these terms should be used, Trost (1981) described a classification system based on why the velopharyngeal mechanism is not functioning as expected. The term “velopharyngeal insufficiency” describes dysfunction resulting from an anatomical or structural defect, as occurs with clefting of the soft palate; whereas, “velopharyngeal incompetence” describes dysfunction resulting from a neuromotor or physiological disorder. The term “velopharyngeal inadequacy” or VPI is suggested for use as the general term for all types of velopharyngeal dysfunction. Despite these distinctions, these terms continue to be used interchangeably in the literature.

The most common cause of velopharyngeal inadequacy is cleft palate (Conley, Gosain, Marks & Larson, 1997). As a result, much of the literature exploring the effects of velopharyngeal inadequacy tends to focus on its effects for speakers with cleft palate. However, VPI may be the result of many other congenital and acquired conditions, including submucous cleft palate, cerebral palsy, coordination problems and facial paralysis (Conley et al., 1997; Peterson-Falzone, 1985).

Closure of the velopharyngeal port is necessary during the production of most speech sounds. When the velopharyngeal mechanism is impaired, both speech and resonance are affected. Hirschberg (1986) described five main characteristics of the speech of individuals with velopharyngeal inadequacy: slow speech development, resonance changes, nasal escape, faulty articulation and hyperfunctional voice production.

The ability of children with VPI to produce and sequence sounds may be affected by their inability to seal off the nasal from the oral cavity during speech. A smaller variety of sounds may be produced by an infant with velopharyngeal inadequacy. As well, utterances may be shorter as greater effort is required to produce speech. Velopharyngeal inadequacy may also affect the functioning of the eustachian tube, leading to frequent middle ear infections and possibly, conductive hearing loss. Hearing loss can also negatively affect speech and language development (Rvachew, Slawinski, Williams & Green, 1999).

Hypernasal resonance is a change in the quality of speech such that it sounds "nasal" (Kummer, 2001, p.157). Although hypernasality is heard predominantly on vowels, inability to close the velopharyngeal port may result in voiced plosives sounding like their nasal counterparts (e.g., "bat" sounds like "mat") (Kummer, 2001).

Nasal air emission or nasal escape refers to the inappropriate release of air through the nasal cavity during speech and is particularly apparent during production of obstruents, as these sounds require a build-up of air pressure in the oral cavity (Kummer, 2001). When air is forced through the velopharyngeal port during speech, a nasal rustle or snort may be heard. Nasal rustles affect the quality and intelligibility of speech (Kummer, 2001). Nasal snorts are most commonly heard in association with the production of sibilants, particularly /s/. As air is being lost through the velopharyngeal opening, consonants may be weak in intensity and pressure or may be omitted completely (McWilliams, Morris & Shelton, 1990). There is an inverse relationship between amount of nasal air emission and the relative strength of obstruent consonant productions. As more air is lost via nasal air emission, obstruent consonants become increasingly weak and, therefore, less distinguishable. To minimize their reliance on velopharyngeal closure during sound production, children with VPI may produce sonorant consonants, which use a more open vocal tract (e.g., /l, r, w, y/) in place of obstruents (Harding & Grunwell, 1996). For example, "pail" may be produced as "whale".

Resonance changes and nasal air escape are considered to be passive speech characteristics (Harding & Grunwell, 1996) or obligatory errors (Trost-Cardamone, 1990). Obligatory errors are not learned but rather are the result of the velopharyngeal abnormality. Compensatory articulation errors, however, are those that are under the individual's control (Trost-Cardamone, 1990). To compensate for VPI, children learn to valve the air stream at places in the vocal tract not commonly used by speakers of English (Gooch, Hardin-Jones, Chapman, Trost-Cardamone & Sussman, 2001). Trost (1981) identified six compensatory articulation errors associated with velopharyngeal inadequacy: glottal stops, pharyngeal fricatives, pharyngeal stops, pharyngeal affricates, mid-dorsum palatal stops and posterior nasal fricatives. Substitution of /h/ for voiceless plosives is another compensatory articulation pattern that may be found in the speech of children with VPI (Harding & Grunwell, 1998). Prevalence studies report that between 22 and 28% of children with cleft palate show compensatory articulation patterns in their speech (Peterson-Falzone, 1990; Dalston, 1990). Prevalence of compensatory articulation patterns for children with VPI without overt cleft palate is unknown.

Hyperfunctional voice disorders may occur in children with velopharyngeal inadequacy because they use the vocal folds more frequently as a place of articulation and constrict the vocal folds together tightly to produce glottal stops. As a result, the child's voice may sound hoarse or rough. Vocal nodules may also develop (Hirschberg, 1986). Vocal nodules have been found in 12% to 15% of children with cleft palate and velopharyngeal inadequacy (Hirschberg, 1986).

Relationship between Speech and Expressive Language

Children with cleft palate often exhibit expressive language delays in addition to articulation or phonological delays. Cleft palate may affect the ability of children to produce and sequence sounds, making speech production more effortful. To compensate for these difficulties and to increase intelligibility, children with cleft palate may shorten the length of their utterances by using primarily content words, resulting in an expressive language delay (Kummer, 2001). Pannbaker (1975) found significant correlations between speech intelligibility and measures of expressive language for adult speakers with cleft palate. As intelligibility decreased, mean length of response and number of words in the longest utterances also decreased.

The majority of research into the relationship between expressive language and articulation has focused on comparing children with and without cleft palate. However, Morris and Ozanne (2003) examined the relationships among the phonetic, phonological and language skills for two groups of three-year-old children with cleft palate. In one group (n = 11), the children had expressive language skills within normal limits. Children in the second group (n = 9) had expressive language skills below age expectations. All children had receptive language skills within normal limits. The comparison of speech and language measures of these two groups of children revealed that the children with delayed expressive language used a smaller number of phonemes productively, used glottal stops as compensatory articulation, and had more cleft-related and developmental error patterns in their speech. The following cleft-related patterns were noted in the speech of the group of children with delayed language: medial consonant deletion, glottal insertion and nasal preference. According to Morris & Ozanne, none of the children with age-appropriate expressive language skills exhibited cleft-related error patterns in their speech.

Speech Characteristics of Children with Syndromic Cleft Palate

Cleft palate is a feature of many syndromes. A syndrome is "a pattern of multiple anomalies that are pathogenically related, and therefore have a common known or suspected cause" (p.79, Kummer, 2001). Kummer (2001) reported that 55% of cases with cleft palate at the Cincinnati Children's Hospital were syndromic or associated with additional anomalies. Teratogens in the environment and/or genetic or chromosomal disorders may result in a syndrome. There are over 370 described syndromes in which cleft palate may occur (Kummer, 2001).

For children with cleft palate and a diagnosed syndrome, speech may be affected by both the cleft and other craniofacial anomalies associated with the syndrome. For syndromes in which craniosynostosis are involved, such as Apert Syndrome, speech may be affected by a cleft palate, a small oral cavity, a narrow palate and hyponasality due to narrowed airways. In other syndromes, such as Stickler syndrome, sensorineural hearing loss is common. Hearing loss may affect one's ability to perceive, discriminate and produce high frequency sounds such as /s/, which may, in turn, affect speech intelligibility. However, the degree to which characteristics of a given syndrome are expressed may vary considerably, making it difficult to characterize the particular speech characteristics of children with cleft palate and a diagnosed syndrome.

Research comparing the speech of children with cleft palate with and without a diagnosed syndrome is limited. Scherer, D'Antonio and Kalbfleisch (1999) found that the language samples of children, aged 30 months, diagnosed with velocardiofacial syndrome (VCFS), with and without an overt cleft, contained significantly more glottal stops, had a significantly smaller consonant inventory and had a significantly lower percent consonants correct (PCC) than peers with cleft lip and palate, with nonsyndromic cleft palate or with typical development (i.e., no history of craniofacial abnormalities). D'Antonio, Scherer, Miller, Kalbfleisch and Bartley (2001) compared the spontaneous speech of children diagnosed with VCFS to children with some of the phenotypic features of the syndrome who did not have the VCFS. In the first group, 92% of children had cleft palate (either overt or submucous) compared to 75% in the second group. They found that the speech of children diagnosed with VCFS who were under seven years old contained a smaller number of consonants, a greater number of developmental errors, a higher frequency of glottal stops and a lower PCC compared to the speech of children without VCFS. These children also seemed to show a preference for voiceless sounds. D'Antonio et al (2001) suggested that the speech production of children with VCFS may be specific to the syndrome, different from children with typical development and from other children with cleft palate.

Children with cleft palate may have additional anomalies but not be diagnosed as having a syndrome. Using five point equal interval rating scales, Persson, Elander, Lohmander-Agerskov and Soderpalm (2000) found that the speech of children with cleft palate and additional anomalies was rated as significantly more hypernasal, and as having significantly weaker pressure consonants, significantly more frequent glottal articulations and significantly more frequent articulation of alveolar plosives as palatal/velar plosives (i.e., /t/ heard as /k/) than the speech of children with cleft palate and no additional malformations. Of the 22 children with additional anomalies, 27% had a diagnosed syndrome (e.g, Stickler syndrome, VCFS). They concluded that children with cleft palate and additional malformations, with or without a diagnosed syndrome, should be considered at risk for severe speech problems.

Speech Intelligibility

Although speech intelligibility is of interest for clinicians involved in the assessment and treatment of children with cleft palate, little research has addressed the question of how to best measure intelligibility in this population (Konst, Weersink-Braks, Rietveld & Peters, 2000). Reliability and validity concerns are often cited as reasons for omitting intelligibility from assessment protocols (Konst et al., 2000). A second reason for omitting this variable is that intelligibility is not affected by cleft-related variables alone (Konst et al., 2000; Wyatt, Sell, Russell, Harding, Harland & Albery, 1986). Prosodic features such as intonation, rate and stress also influence the intelligibility of cleft palate speakers (Wyatt et al., 1986), making it more difficult to distinguish the relative contribution of suprasegmental features and cleft-related perceptual features to reduced speech intelligibility.

The relationship between intelligibility and articulation in cleft palate speech is also unclear (Subtelny, Van Hattum, & Myers, 1972). McWilliams (1954) found a significant positive correlation (r = .72) between intelligibility as determined by orthographic transcriptions and articulation scores, whereas Subtelny et al. (1972) found poor agreement between intelligibility and articulation as rated on two six-point scales.

Subtelney et al. (1972) also found that speakers could have a severe articulation disorder yet still be rated as intelligible.

Whitehill and Chun (2002) examined the relationship between intelligibility and articulatory accuracy, as measured by the percentage of phonemes phonetically transcribed as correct. Intelligibility was based on children's single word productions and was evaluated using a closed-set response task format. Intelligibility was defined as the percentage of words identified correctly. A significant positive correlation ($r = .77$, $p < .01$) was found between intelligibility and articulatory accuracy scores.

Hypernasality is a common feature of children with cleft palate. Keuning, Wieneke, van Wijngaarden and Dejonckere (2002) examined the relationship between nasality and intelligibility for speakers with cleft palate. Nasality was assessed by visual analog scale and nasalance measures obtained using a nasometer. Intelligibility was evaluated using a visual analog scale. Anchors on both scales were defined as "normal" (left side) and "extremely deviant" (right side). The perceptual ratings of nasality and intelligibility had positive correlations of $r = .63$ for material containing nasals and $r = .60$ for samples without nasals, such that samples rated as more deviant in terms of hypernasality were also rated as more deviant in terms of intelligibility. Although significant correlations were also found between intelligibility ratings and nasalance scores for both types of stimuli, these correlations were quite low (with nasals: $r = .37$ and without nasals: $r = .34$). Conversely, Whitehill and Chun (2002) did not find a significant relationship between intelligibility, determined using a closed-set task, and nasality, as rated on an equal interval scale. The differences in correlations between intelligibility and nasality in these studies may be related to differences in measurement procedures and/or to differences in composition of the groups. Whitehill and Chun (2002) examined data from mildly affected children aged 5 to 15 years; whereas, Keuning et al. (2002) collected data from subjects aged 4 to 83 years with a wide range of severity. It is expected that hypernasality would have a greater effect on intelligibility when it is more severe, as it may cause oral consonants to sound like nasal ones (e.g., "bad" to "mad"). Maegawa, Sells and David (1998) reported that subjects with cleft palate and reduced intelligibility had increased hypernasality, compared to subjects with cleft palate and improved intelligibility, in their investigation of speech changes after maxillary advancement.

Whitehill (2002) conducted a review of intelligibility studies of cleft palate speakers for the period of 1960 – 1998. This review revealed that over 20% of the studies did not adequately define or differentiate the terms intelligibility, acceptability, naturalness, severity and proficiency. Furthermore, almost 50% of the studies reviewed used interval scaling to measure intelligibility. A further 16% used a crude estimation of intelligibility. Both interval scaling and estimation are not considered valid methods for measuring speech intelligibility (Schiavetti, 1992; Gordon-Brannan, 1994). Whitehill (2002) recommended evaluating speech intelligibility using measures that are both reliable and valid, such as transcription tasks (i.e., open-set response task), multiple-choice tasks (i.e., closed-set response task) and magnitude estimation tasks. Transcription and multiple-choice tasks also have the advantage of making it easier for listeners to focus solely on intelligibility (Whitehill, 2002).

Whitehill and Chau (2004) developed a single-word intelligibility test for Cantonese speakers with cleft palate and reported results for a group of 15 speakers with cleft palate who ranged in age from 5 – 44 years (*M* = 24.5 years). This test uses a closed-set response task format (four choices) and targets phonetic contrasts known to be problematic for speakers with cleft palate. The percentage of items identified correctly by listeners in the closed-set response task served as the intelligibility score. Scores ranged from 32.3% to 86.7% for the 15 speakers.

Speech Intelligibility Probe for Children with Cleft Palate Version 2

To date, there are no published tools to measure the speech intelligibility of children with cleft palate. The Speech Intelligibility Probe for Children with Cleft Palate Version 2 (SIP-CCLP Ver. 2) is a computer-administered word imitation measure that is under development at the University of Alberta. The SIP-CCLP Ver. 2 is one of a set of measures of speech intelligibility being developed as part of the Test of Children's Speech Plus (TOCS+) project (www.tocs.plus.ualberta.ca). It has been designed to measure speech intelligibility, using both open (i.e., orthographic transcription) and closed-set (i.e., multiple choice) word identification tasks, and to be sensitive to the error patterns found in the speech of children with cleft palate. Version 2 was developed from the Children's Intelligibility Probe for Cleft Palate (CIP-CLP) (Connolly, 2001) and SIP-CCLP Ver. 1 (Feltz et al., 2002). This tool uses a phonetic contrast approach to intelligibility assessment (Connolly, 2001; Kent et al., 1989). Word pairs that vary systematically in their consonant constituents and that target the error patterns of interest for children with cleft palate serve as stimuli.

Development of SIP-CCLP Ver. 2

Connolly (2001) developed the original stimulus words for the Children's Intelligibility Probe for Cleft Palate and tested it with one child with cleft palate. A discussion on how stimulus words were selected is included in the following section. The Children's Intelligibility Probe for Cleft Palate used a flashcard format with clipart pictures to elicit word productions. Feltz et al. (2002) revised this tool to include items to sample consonant cluster reduction, replaced clipart with original artwork, incorporated the stimulus items into a slide presentation using Microsoft PowerPoint (Microsoft, 2000) and renamed it the Speech Intelligibility Probe for Children with Cleft Palate (SIP-CCLP Ver. 1). For both the Children's Intelligibility Probe for Cleft Palate and SIP-CCLP Ver. 1, children's word productions were recorded using an analog audio-video camera and converted to audio digital recordings for playback using CSpeech 4.0 (Milenkovic & Read, 1994).

Listener judgments for both the Children's Intelligibility Probe for Cleft Palate and SIP-CCLP Ver. 1 were obtained using two paper and pencil tasks. In the first task (open-set response task), listeners wrote down what they heard the child say and the examiner later compared these orthographic transcriptions with a key to determine percent words identified correctly. In the second task (closed-set response task), listeners chose which of three words (the target and two minimally contrastive words) or "? (other)" best matched with what was heard. Then, they rated how confident they were in their choice. From the listeners' responses, information was obtained about the child's intelligibility and the type of errors in the child's speech.

Feltz et al. (2002) conducted a preliminary evaluation of the validity (content, construct, and criterion) and reliability (interjudge, intrajudge) of SIP-CCLP Ver. 1 using data from five children with typical speech and language development. Their results suggested that the SIP-CCLP Ver. 1 had potential to be a valid and reliable clinical tool.

Feltz et al. (2002) recommended that the validity and reliability of SIP-CCLP Ver. 1 be evaluated further using a larger sample of children, including children with and without cleft palate.

Data were also collected by Feltz et al. (2002) for two children with cleft palate but not analyzed. Gotzke & Hodge (2003) compared these data with the results from children with typical speech development reported by Feltz et al. (2002). One child from the Feltz et al. (2002) study was excluded as it was determined that the recording levels for SIP-CCLP Ver. 1 were two low for valid listener judgments. Consequently, the results from the two children with cleft palate were compared to the remaining four subjects with typical speech from Feltz et al. (2002). Results of this analysis suggested that SIP-CCLP Ver. 1 samples the kinds of speech errors made by children with cleft palate and is also sensitive to developmental errors.

As a result of limitations identified by Gotzke and Hodge (2003) and recommendations made by Feltz et al. (2002), Gotzke (2003) revised the presentation format, stimulus word items, phonetic contrast items, confidence rating procedure and speech acceptability rating scale of the SIP-CCLP Ver. 1 and renamed the tool, SIP-CCLP Ver. 2. Additional artwork was developed for the new stimulus words and all pictures were enhanced using Adobe Photo Deluxe (Bailey et al., 1989) to increase their appeal to children. Software, based on that developed for the Test of Children's Speech + (TOCS+) (www.tocs.plus.ualberta.ca), was created using Authorware 6.0 (Macromedia, 2001) to enable SIP-CCLP Ver. 2 to be computer administered. The SIP-CCLP Ver. 2 software program randomly generates the presentation order of the target words and allows children's word productions to be audio-recorded directly to the computer hard drive as .wav files. As well, the software has new features to administer the listening tasks and save each listener's responses in computer files for scoring and analysis. These revisions to SIP-CCLP Ver. 1 are described in greater detail in Gotzke (2003). Gotzke (2003) also evaluated the validity (construct, criterion) and reliability (interjudge, intrajudge) of SIP-CCLP Ver. 2. Tables 1 and 2 show the developmental history of the SIP-CCLP Ver. 2 features.

Table 1.

Development of SIP-CCLP Ver. 2 Administration: Child Recordings

	Tool Name	Presentation	# Stimulus Items	Recording Mode	Number of Subjects	
					Typical	Cleft
Connolly (2001)	Children's Intelligibility Probe for Cleft Palate	Cards	107	Analog	0	1
Feltz et al. (2002)	SIP-CCLP Ver.1	PowerPoint	118	Analog	5	0
Gotzke & Hodge (2003)	SIP-CCLP Ver. 1	PowerPoint	118	Analog	5[1]	2
Gotzke (2003)	SIP-CCLP Ver.2	Custom Software Program	123	Direct Digital Record	8[2]	4[3]

Note. [1]Data from Feltz et al. (2002); [2]Includes four children from Feltz et al. (2002); [3]Includes two children from Gotzke & Hodge (2003).

Table 2.

Development of SIP-CCLP Ver. 2 Administration: Listener Response Tasks

	Tool Name	Response Format	# Reliability Items (Open-set Response Task)	# Items in Closed-set Response Task	Other Ratings in Closed-set Response Task
Connolly (2001)	Children's Intelligibility Probe for Cleft Palate	Paper/Pencil	0	160	Confidence (very/somewhat/ not)
Feltz et al. (2002)	SIP-CCLP Ver. 1	Paper/Pencil	12	171	Confidence (very/somewhat/ not)
Gotzke & Hodge (2003)	SIP-CCLP Ver. 1	Paper/Pencil	12	171	Confidence (very/somewhat/ not)
Gotzke (2003)	SIP-CCLP Ver.2	Computer Screen/ Mouse, Keyboard	12	194	Distortion (clear/distorted)

Note. There were no reliability items for the closed-set response task in these versions of SIP-CCLP.

Description of SIP-CCLP Ver. 2 Contrast Pair Development

As SIP-CCLP Ver. 2 uses a phonetic contrast approach to intelligibility assessment, a first step in its development was to determine what error patterns are commonly found in the speech of children with cleft palate and velopharyngeal inadequacy (Connolly, 2001; Feltz et al., 2002). The results of a literature review revealed six different types of errors: manner preferences (MP), place preferences (PP), glottal errors (GE), sibilant errors (SE), voicing errors (VE), and cluster errors (CE). A manner preference occurs when a child consistently uses a sound with a different manner of articulation than the target sound (e.g., stop to glide: "pail" to "whale"). Place preferences occur when a sound is articulated at a different place of articulation than the target sound (e.g., alveolar to bilabial: “tail” to “pail”). Glottal errors include instances in which a glottal stop (i.e., /ʔ/) or glottal fricative /h/ is used in place of some other consonant (e.g., alveolar to glottal: “cat” to “hat”). Sibilant errors are instances in which a sibilant (i.e., /s/, /z/, /ʃ/ or /ʒ/) is produced at a different place of articulation (e.g., alveolar to palatoalveolar: “sip” to “ship”). Voicing errors occur when a voiced consonant is used in place of its voiceless counterpart or vice versa (e.g., voiceless to voiced: “rope” to “robe”). Cluster errors occur when a consonant(s) is deleted from a cluster (e.g., “stop” to “top”).

Cleft-related errors.

Within the six major categories defined above, the following error patterns associated with cleft palate were identified: place preferences for velars (e.g., alveolar to velar: “down” to “gown”), place preference for bilabials (e.g., alveolar to bilabial: “tail” to “pail”), sibilant error palatalization (e.g., alveolar to palatoalveolar: “sip” to “ship”), sibilant error weakening (e.g., alveolar to interdental: “sat” to “fat”), manner preference for liquids (e.g., stop to glide: “bays” to “ways”), manner preference for glides (e.g., fricative to glide: “shell” to “yell”), manner preference for nasals (e.g., stop to nasal: “D” to “knee”), glottal error substitution of glottal stop (e.g., alveolar to glottal: “toe” to “ʔO”), glottal error substitution of glottal fricative (e.g., alveolar to glottal: “tail” to “hail”) and voicing error in final position (e.g., voiceless to voiced: “rope” to “robe”). These errors are thought to be associated with the presence of fistulas, structural differences at the alveolar ridge and/or velopharyngeal inadequacy.

Developmental errors.

The literature review also revealed developmental error patterns that have been found in the speech of children with cleft palate. These sounds require more precise control of the tongue or increased coordination of the muscles of the vocal tract with the oral articulators and as a result, are often mastered later. Children may substitute sounds they can already produce for those they are learning, resulting in what are described as "developmental error patterns," in the speech of children with typical speech and language development. Opportunities for making the following developmental error patterns were included in SIP-CCLP Ver.2: manner preference for stops (e.g., fricative to stop: "sail" to "tail"), place preference for alveolars (e.g., velars to alveolars: "gown" to "down"), sibilant error fronting (e.g., palatoalveolar to alveolar: "shell" to "sell"), word initial voicing errors (e.g., voiceless to voiced: "fail" to "veil") and cluster reduction errors (e.g., "stop" to "top"). For children without cleft palate, only developmental errors are expected; whereas, for children with cleft palate, both cleft-related and developmental errors are expected (Harding & Grunwell, 1996).

Unknown errors.

A third error category, called "unknown" was added during development of SIP-CCLP Ver.2. This category includes errors for which there is insufficient information to determine if the pattern is cleft-related. For example, there is conflicting information in the literature regarding the nature of voicing errors for children with cleft palate. McWilliams, Morris & Shelton (1990) reported that these children make more errors on voiceless than voiced sounds, whereas, O'Gara & Logemann (1988) stated that children with cleft palate produce voiceless sounds earlier than voiced. Because of these differences, voicing errors where voiceless sounds are substituted for voiced sounds (e.g., "tear" for "dear") were included in the unknown category. Opportunities to sample the remaining error patterns in the unknown category were included to allow both members of each contrast pair to be used as targets and to test whether other errors not attributed to cleft palate or found during normal development may be present. An example of an unknown error pattern would be a manner preference where a stop is substituted for a liquid (e.g., "low" to "toe"). Appendix A contains a list of all stimulus words used in SIP-CCLP Ver. 2. Target words within contrast word pairs organized by classification as cleft-related, developmental or unknown and type of error pattern are included in Appendix B.

Description of SIP-CCLP Ver. 2 Administration

SIP-CCLP Ver. 2 administration consists of two parts: recording a child's word productions and playing these back to listener judges. For each child, the SIP-CCLP software randomly generates the presentation order of the 123 SIP-CCLP Ver. 2 stimulus words. Each word corresponds with a picture shown on the computer screen. The child's word productions are elicited using an imitative format and are directly recorded to the computer as digital audio files. The actual stimulus words are preceded by four practice items. The SIP-CCLP Ver. 2 software allows the examiner to repeat the practice items to ensure the child understands the task. After recording each stimulus word, the examiner can playback the child's productions to check recording quality and elicit the child's production a second time if he/she has concerns about background noise or examiner voiceover. An example of a stimulus word presentation screen is provided in Appendix

C. Short breaks are provided in the form of musical computer animations that appear after every 20 items.

Listener judges complete two different response tasks on the computer. In the open-set response task, each listener is instructed to use the keyboard to orthographically transcribe what he/she hears. The listener is instructed to type in the sound(s) heard if the word can not be identified. From the child's word productions, the software randomly selects twelve of the stimulus words for repeat playback. Repeated items are used to determine intrajudge reliability by calculating number of agreements over the two presentations. An example of the open-set response task computer screen, including listener judge instructions, can be found in Appendix D. The software saves the listeners' responses to files for later analysis.

For the closed-set response task, the listener is shown four choices ("buttons") on the computer screen. The listener judge is instructed to click the mouse on the button that best matches the sound heard in the target position. The target sound is highlighted using a different color font and underlining. The first two buttons are the minimal pair contrast items. For example, if the error pattern of place preference for velars was being tested, the two buttons would be "cape" and "tape" presented in alphabetical order with "c" and "t" underlined and in a different color font. The third button is a "blank" in which the listener can use the keyboard to type in the sound that was heard if it differs from the highlighted target in the first two buttons. The fourth button is labeled "can't identify" and is intended for use in situations where what was heard cannot be identified as an English sound. If the listener judge chooses buttons 1, 2 or 3, a second box appears on the screen. The listener judge then indicates via mouse click whether the sound production in the target location is "distorted" or "clear". Appendix E provides an example of the closed-set response task instruction screen. Appendix F provides an example of the closed-set response task presentation screen for choosing a button and then rating whether the production as distorted or clear. Listener responses are saved to files for later analysis. Following completion of the open-set and closed-set response tasks in SIP-CCLP Ver. 2, listeners also rate the overall acceptability of the child's speech and then, the overall amount of effort that was required to understand the child's speech (listener effort) using two different five point equal interval scales.

The percentage of words identified correctly in the open-set response task provides an intelligibility score for the child. Intrajudge reliability for the open-set response task is determined by comparing judges' responses on the first and second presentation of the twelve repeated items and determining the number of agreements. The analysis of the closed-set responses provides information on the type of error patterns that are present in the child's speech and the frequency of sounds identified as being in error. To determine whether an error pattern is present in the child's speech, a minimum of two out of three listeners must choose or type in the same response. Distortion ratings are also calculated based on listener responses in the closed-set response task. These ratings provide another measure of speech sound accuracy and a way to capture the distortions often present in the speech of children with cleft palate.

Validity and Reliability of SIP-CCLP Ver. 2

Gotzke (2003) reported results for a preliminary evaluation of the validity (construct and criterion) and reliability (interjudge, intrajudge) of the SIP-CCLP Ver. 2. Results for eight children with typical speech and language development and no history of cleft lip and/or palate and four children with cleft palate and normal language development were analyzed. This analysis included data for four children with typical speech development from Feltz et al. (2002) and two children with cleft palate from Gotzke and Hodge (2003). Each group had a mean age of 57 months. Table 3 provides descriptive information for each child participant. Each child's SIP-CCLP Ver. 2 word productions and a 100-word sample of spontaneous speech were heard by three independent listener judges. Listener judges were 35 students in speech-language pathology. All had English as a first language and normal hearing. Results of the evaluation of content validity by Feltz et al. (2002) and of criterion and construct validity and reliability by Gotzke (2003) are described in the following sections.

Content validity.

Content validity is established when experts judge that the items on a measure are representative of the phenomenon being measured (Vogt, 1993). Content validity was considered during development of the Children's Intelligibility Probe for Cleft Palate (Connolly, 2001) and evaluated by Feltz et al. (2002). In both studies, speech-language pathologists with clinical expertise working with children with cleft palate found that the SIP-CCLP Ver. 1 items adequately sampled the characteristics they purported to measure. A concern expressed by one clinical expert was that it would be difficult to capture the nonstandard substitutions and distortions that are common in cleft palate speech using SIP-CCLP Ver. 1 (Feltz et al., 2002). Gotzke (2003) attempted to address this concern in SIP-CCLP Ver. 2 by having listeners rate whether the child's production of the target was distorted (=2) or clear (=1) after making their response. As well, listeners were able to type in a response if it did not match one of the two choices provided. In this blank, listeners could also type in an abbreviation for the compensatory productions (i.e., glottal stops, pharyngeal fricative, pharyngeal stop, pharyngeal affricates, mid-dorsum palatal stops and posterior nasal fricatives) described by Trost (1981) if desired. The apriori classification of SIP-CCLP Ver. 2 error patterns into cleft-related and developmental was supported by Morris and Ozanne (2003). Overall, clinical experts concluded that SIP-CCLP Ver. 1 items had content validity.

Table 3.

Child Participant Characteristics for Gotzke (2003)

Subject Group	Subject Number	Age in Months	Gender	Description of Cleft
Typical Speech	TS1[1]	37	female	n/a
	TS2[3]	43	female	n/a
	TS3[3]	57	male	n/a
	TS4[3]	57	male	n/a
	TS5[3]	60	male	n/a
	TS6[1]	61	male	n/a
	TS7[1]	65	male	n/a
	TS8[1]	82	male	n/a
Cleft Palate	CP1[3]	45	female	Bilateral cleft lip and palate (repaired)
	CP2[3]	47	male	Unilateral cleft lip and palate (repaired)
	CP3[2]	59	male	Unilateral cleft lip and palate (repaired)
	CP4[2]	79	male	Unilateral cleft lip and palate (repaired)

Note. [1]Data from Feltz et al (2002); [2]Data from Gotzke and Hodge (2003); [3]Data from Gotzke (2003).

Construct validity.

Construct validity is the degree to which a measure reflects some theoretical explanation of a behavior or characteristic being measured (Schiavetti & Metz, 2002). Construct validity of the SIP-CCLP Ver .2 was assessed by Gotzke (2003) with respect to predictions about a) intelligibility scores, b) error patterns, c) distortion ratings and d) speech intelligibility score correlations with listener ratings of effort and acceptability.

Gotzke's (2003) prediction that the mean intelligibility score for the children with cleft palate would be significantly lower than the mean intelligibility score for the children with typical speech development on both the open-set and closed-set response tasks of the SIP-CCLP Ver. 2 was supported. The mean intelligibility scores for the children with and without cleft palate on the open-set response task were 50.15% and 72.07%, respectively. The mean intelligibility scores on the closed-set response task were 82.05% (cleft palate) and 94.0% (typical speech). Individual subject and group scores from Gotzke (2003) are summarized in Table 4.

Gotzke (2003) predicted that while both developmental and cleft-type error patterns should be present in the speech of children with cleft palate, cleft-type errors should predominate. As predicted, more cleft-related (47.4%) than developmental errors (3.1%) were identified for the children with cleft palate. This result suggests that the SIP-CCLP Ver.2 is sampling the kinds of error patterns commonly seen in the speech of children with cleft palate. Gotzke (2003) also predicted that for children without cleft palate, developmental error patterns should predominate. This hypothesis was supported as developmental errors (23.2%) were more frequent than cleft-related errors (15.9%) for the children with typical speech development.

Overall, the mean percentage of errors for the children with and without cleft palate was 17.85% and 5.65%, respectively. These results are similar to those reported by Chapman (1993) in which children, aged three to five years, with and without cleft palate were compared in terms of frequency of phonological errors using the Goldman-

Fristoe Test of Articulation (Goldman & Fristoe, 1986) and Khan-Lewis Phonological Analysis (Khan & Lewis, 1986). In the Chapman (1993) study, as in Gotzke (2003), the speech of children with cleft palate contained more phonological errors than the children without cleft palate.

Gotzke (2003) predicted that children with cleft palate should have a significantly higher percentage of obstruent than sonorant errors, whereas children without cleft palate should exhibit no significant difference between percentage of obstruent and sonorant errors. As predicted, the children with cleft palate had a greater percentage of obstruent than sonorant errors (21.9% and 9.54%, respectively). However, this result was not significant ($t = 1.99$, $p > .05$), possibly due to the large variability in percentage of errors among the children with cleft palate. Results confirmed the prediction that children with cleft palate would exhibit a significantly greater percentage of obstruent errors than children with typical speech development. Percentage of obstruent errors for children with and without cleft palate was 21.9% and 6.99%, respectively. This trend is in keeping with previous studies in which children with cleft palate were found to make more errors on obstruents (Van Demark, Morris & Van De Haar, 1979), suggesting that SIP-CCLP Ver. 2 is sampling the sound classes commonly in error in the speech of children with cleft palate.

Table 4.

SIP-CCLP Ver. 2 and spontaneous speech sample intelligibility, and SIP-CCLP Ver. 2 closed-set analysis error scores for children without and with cleft palate

Subject Number (Age in months)	Mean Intelligibility Score (%)			SIP-CCLP Closed-set Analysis Error Scores (%)		
	SIP-CCLP: Open-set	SIP-CCLP: Closed-set	Spontaneous Speech: Open-set	Mean Obstruent Errors	Mean Sonorant Errors	Distorted Items
TS1 (37)[1]	52.26	87.33	78.30	16.01	7.27	24.37
TS2 (43)[3]	88.89	98.79	82.18	1.41	0	10.54
TS3 (57)[3]	81.7	95.34	94.90	4.93	0	3.63
TS4 (57)[3]	73.71	85.32	86.4	17.46	11.97	7.77
TS5 (60)[3]	88.8	98.1	90.66	1.18	0.52	3.63
TS6 (61)[1]	74.58	95.91	83.9	4.72	3.04	12.09
TS7 (65)[1]	68.93	93.37	87.20	8.4	2.42	15.01
TS8 (82)[1]	89.27	97.86	95.90	1.84	2.42	4.29
Typ. Group (*M, SD*)	77.27 (12.8)	94.0 (5.08)	87.43 (6.12)	6.99 (6.48)	3.45 (4.17)	11.13 (6.61)
CP1 (45)[3]	34.96	73.4	47.00	30.75	13.76	36.44
CP2 (47)[3]	51.49	81.52	55.00	21.16	17.13	32.64
CP3 (59)[2]	35.88	76.41	56.60	32.54	4.86	26.51
CP4 (79)[2]	78.25	96.88	92.20	3.15	2.42	7.21
Cleft Group (*M, SD*)	50.15 (20.21)	82.05 (10.44)	62.7 (20.11)	21.90 (13.46)	9.54 (7.03)	25.7 (12.99)

Note. [1]Data from Feltz et al (2002); [2]Data from Gotzke and Hodge (2003); [3]Data from Gotzke (2003)

As sounds produced by children with cleft palate are often judged as distorted, Gotzke (2003) predicted that children with cleft palate would have a higher percentage of items rated as distorted and a higher percentage of "correct distorted ratings" than children without cleft palate. The mean percentage of items rated as distorted was 25.7% for the children with cleft palate and 11.13% for the children with typical speech development. The mean percentage of "correct distorted ratings" for the children with and without cleft palate was 21.12% and 9.2% respectively. These results provide support for both of the hypotheses, suggesting that using SIP-CCLP Ver.2, listeners are able to identify sound distortions present in the speech of children with cleft palate.

Finally, Gotzke (2003) predicted positive relationships between acceptability and intelligibility (i.e., greater acceptability, greater intelligibility of speech) and a negative relationship between listener ratings of effort and intelligibility (i.e., the lower the intelligibility score on the open-set response task, the higher listeners rated the amount of effort required to understand what the child had said). Both of these relationships were statistically significant (acceptability-intelligibility: $r = 0.82$, $p<.05$, effort-intelligibility: $r = -0.73$, $p=.05$). The positive relationship found between acceptability and intelligibility in this study agrees with results from Whitehill and Chun (2002). Gotzke (2003) also predicted that the median listener acceptability ratings would be lower for the children with cleft palate than the children with typical speech, whereas the median listener effort ratings would be higher for the children with cleft palate than children without cleft palate. Neither of these hypotheses was supported. This result may be explained by the small number of subjects and the variability in severity of speech disorder in the children with cleft palate.

This evaluation of construct validity suggests that SIP-CCLP Ver.2 is effective at identifying the overall intelligibility of children with cleft palate, the types of error patterns present in their speech, and the sound types in error. However, Gotzke (2003) recommended further evaluation of the construct validity of this tool with a larger sample of children with and without cleft palate before implementing it in clinical practice.

Criterion validity.

Criterion validity can be defined as the ability of a tool to make accurate predictions and is measured by determining how well it predicts an outside criterion (Schiavetti & Metz, 2002). To evaluate the criterion validity of SIP-CCLP Ver. 2, the degree to which intelligibility scores on the SIP-CCLP Ver. 2 open-set response task correlated with intelligibility scores obtained from an analysis of a sample of the child's spontaneous speech was determined for the eight subjects with typical speech and four children with cleft palate described in Table 3 (Gotzke, 2003). Analysis of the child's spontaneous speech followed the procedures described by Weiss (1980). A statistically significant positive correlation ($r = 0.89$) was found between the two intelligibility scores for the twelve subjects. This result agrees with results from Gordon-Brannan and Hodson (2000) who reported a positive correlation ($r = 0.79$) between intelligibility scores for spontaneous speech and imitated single words for 48 pre-kindergarten children with varying phonological abilities. This result suggests that SIP-CCLP Ver. 2 has criterion validity.

Criterion validity was also evaluated by comparing the results of phonetic analysis of the SIP-CCLP Ver. 2 stimulus words and the spontaneous speech sample for all 12 children. No significant difference was found in the percentage of errors for stops, fricatives, affricates, nasals and glides between these two sampling conditions, supporting the criterion validity of SIP-CCLP Ver. 2. However, liquids were transcribed as correct significantly more often in the spontaneous speech sample. The results of the phonetic analysis of the spontaneous speech sample were also compared with an error analysis of the SIP-CCLP Ver. 2 closed-set responses. No significant differences were found between the two sampling conditions for all manner classes, except stops. A significantly greater percentage of stops were identified as correct in the closed-set response task than were transcribed as correct in the spontaneous speech sample. In a study comparing the results of articulation testing and spontaneous speech sampling for children with a speech delay, Morrison and Shriberg (1992) found that the percentage correct for nasals and glides was higher in a spontaneous sample, whereas the percentage correct for affricates and fricatives were higher in articulation testing (i.e., single-word citation). Percentage correct for stops did not differ between the two sampling conditions (Morrison & Shriberg, 1992). This difference in results between Morrison and Shriberg (1992) and Gotzke (2003) suggests that SIP-CCLP Ver. 2 is more effective in obtaining a representative sample of the everyday speech of children with cleft palate.

Gotzke (2003) concluded that SIP-CCLP Ver. 2 satisfied the standards for criterion validity because a high positive correlation was found between intelligibility scores obtained from the SIP-CCLP Ver. 2 open-set response task and the spontaneous speech sample using word identification tasks. With the exception of liquids, no significant difference between sampling condition was found when phonetic transcriptions of the SIP-CCLP Ver. 2 words and spontaneous sample were compared. In addition, no significant difference in percentage of errors was found for all manner classes, except stops, when closed-set task responses were compared with phonetic transcription of the spontaneous speech sample. Further examination of the criterion validity of SIP-CCLP Ver. 2 was deemed necessary to determine if the preliminary results reported by Gotzke (2003) held for a larger sample of children, including children with cleft palate.

Reliability.

Gotzke (2003) also assessed the reliability of SIP-CCLP Ver. 2 for the eight children with typical speech and four children with cleft palate described in Table 3. Interjudge reliability on the open-set response task of SIP-CCLP Ver. 2, based on sets of three listeners per child for a total of 36 listeners, was high (intraclass correlation coefficient or ICC = 0.98), suggesting that intelligibility can be assessed reliably using the SIP-CCLP Ver. 2 single words. Mean intrajudge agreement on the twelve repeated words in the open-set response task was 82.4%. This value is comparable to results from other studies in which open-set word identification tasks were used (Hodge, 1996). Intrajudge reliability was not assessed with respect to the SIP-CCLP Ver. 2 closed-set response task.

Recommended Changes to Revise SIP-CCLP Ver. 2 to Version 3

The results of these assessments of the validity and reliability of SIP-CCLP Ver. 2 were based on sample of eight children with typical speech development and four children with cleft palate. Further evaluation of the validity and reliability of SIP-CCLP Ver. 2 using a larger sample of children with and without cleft palate appeared warranted.

Based on a review of the literature and recommendations made by Gotzke (2003), the SIP-CCLP Ver. 2 software was revised. Recommendations for SIP-CCLP Ver. 3

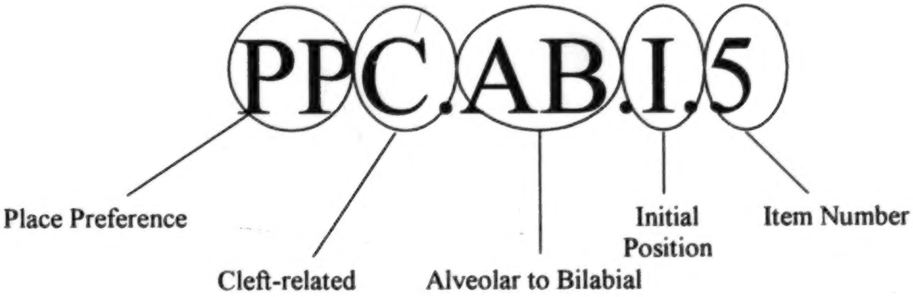
included revising the error coding of the target phonetic contrast items for the closed-set response task, creating analysis software and improving the error profiling and graphing capabilities of the software. As well, a feature to determine intra-rater reliability on the closed-set response task was added. Each of these recommended revisions is elaborated in the following sections.

Analysis of the listener responses for SIP-CCLP Ver. 2 revealed that the target word "fell" was not elicited from children. As a result, the cleft-related manner preference error "fell - yell" was not tested in the closed-set response task. For SIP-CCLP Ver. 3, it was recommended that this target word be added to the SIP-CCLP Ver. 2 stimulus words and the target phonetic contrast item be added to the SIP-CCLP Ver. 2 closed-set response task. These changes increased the number of stimulus words to 124 and the number of phonetic contrast items to 194.

SIP-CCLP Ver.3 closed-set error coding.

The error coding system of SIP-CCLP Ver. 2 was revised to better reflect the error patterns that were identified in the literature review. In the new error coding scheme, each pattern is first classified as to whether it is a manner preference (MP), place preference (PP), glottal error (GE), sibilant error (SE), voicing error (VE) or cluster error (CE). The third position in the code identifies whether the error pattern is cleft related (C), developmental (D) or unknown (U). The fourth and fifth position in the code contain information about the direction of sound change with the code for the target sound listed first, followed by the code for the foil. For example, with the error pattern PPC.AB.I.5 where the target and foil are "tail" and "pail," respectively, the sound change is from alveolar (A) to bilabial (B). The sixth position in the code indicates whether the target sound is in initial (I), medial (M) or final (F) position in the word. The final position in the code is a unique numerical identifier for that target sound contrast. Figure 1 shows a breakdown of the code for an example target contrast pair "tail" and "pail." Appendix B contains a full list of all the new codes for the phonetic contrast pairs for SIP-CCLP Ver. 3. These codes better reflect the kinds of error patterns that are tapped in SIP-CCLP Ver. 3, and thereby improve the profiling capabilities of the software.

Figure 1. Example of error pattern coding for SIP-CCLP Ver. 3 for the phonetic contrast pair "tail" (target) and "pail" (foil).



SIP-CCLP Ver. 3 closed-set analysis software

It was recommended that an analysis component that allows the user to compile listener responses for the closed-set response task be added to SIP-CCLP Ver. 3. It was envisioned that the analysis software would first ask the user whether he/she would like to examine listener responses for the items in which a response was typed in the "blank" button. By examining the "blank" responses, the user could recode instances in which the listeners typed in the same sound as one of the two choices provided or typed in a word that contained a sound other than the target sound in error. For example with the error pattern PPC.AB.I.5, in which listeners hear the sound file for "pail", listeners would be provided with the two choices: "pail" and tail." If the listener typed in "pain" in the blank button, the analysis software will allow the user to change the scoring to identify that item as correct, as the target sound ("p") was identified as being produced correctly.

Once the "blank" buttons had been examined, the analysis software could then compile the closed-set response task responses for three listeners. The software would then identify items in which at least two of three listeners chose the same response (i.e., target, foil or "can't identify") or typed the same response in the "blank." These target phonetic contrast items could then be analyzed further to determine the type of error pattern represented and the sound in error. Items for which no consensus is reached (i.e., each judge chose a different response) would be marked incorrect but not used in the analysis of error patterns. As well, the number of correct and distorted ratings could be totaled for each group of listeners. These recommended changes to the analysis software are yet to be completed.

SIP-CCLP Ver. 3 closed-set error profiling and graphing.

The recommendations for developing the error profiling and graphing software were not completed for this assessment of validity and reliability but are recommended for the final version of SIP-CCLP. Error profiling and graphing were done manually following the format recommended for the software features.

SIP-CCLP Ver. 3 closed-set response task reliability.

In SIP-CCLP Ver. 2, intrajudge reliability was determined for the open-set response task only. SIP-CCLP Ver. 3 evaluates intrajudge reliability for both the open-set and closed-set response tasks. To evaluate intrajudge reliability of the listener judges' closed-set responses, the SIP-CCLP Ver. 3 software randomly chooses 20 minimal contrast items to be repeated. Listener responses on the first and second presentation of these items are then compared to determine agreement.

Summary.

The error coding of SIP-CCLP Ver. 3 target phonetic contrast items was revised. This change complements the development of closed-set analysis software that will allow "blank" responses to be analyzed and results to be profiled and graphed. As well, repeated items were added to allow assessment of intrajudge reliability on the closed-set response task.

Purpose

The purpose of this study was to further evaluate the validity (construct, criterion) and reliability (interjudge, intrajudge) of SIP-CCLP Ver. 3 as a measure of speech intelligibility by collecting and comparing data from a sample of children with and without cleft palate and by comparing data from a sample of children with typical speech development of different ages.

Research Questions and Hypotheses

Reliability

With respect to interjudge reliability, the following questions were asked: 1) What is the interjudge reliability on the SIP-CCLP Ver. 3 open-set response task? 2) What is the interjudge reliability on the identification of items as "correct/incorrect" and judgments of items as "clear/distorted" on the SIP-CCLP Ver. 3 closed-set response task?

Specifically, the following hypotheses were tested:

1. A strong positive correlation is expected for interjudge reliability (ICC >0.8) on the SIP-CCLP Ver. 3 open-set response task.
2. Interjudge agreement on the identifications of "correct/incorrect" and judgments of "clear/distorted" on the SIP-CCLP Ver. 3 closed-set response task will be greater than 80%.

With respect to intrajudge reliability, the following questions were asked: 1) What is the intrajudge reliability for identification of "correct/incorrect" on the SIP-CCLP Ver. 3 open-set and closed-set response tasks? 2) What is the intrajudge reliability for judgments of "clear/distorted" on the SIP-CCLP Ver. 3 closed-set response task?

1. Based on previous findings from Gotzke (2003) and Hodge (1996), it was expected that the mean intrajudge agreement would be greater than 80% and positively related to the child's intelligibility score for identification of "correct/incorrect" on both SIP-CCLP Ver. 3 response tasks.
2. It was also predicted that mean intrajudge agreement on the judgments of "clear/distorted" on the SIP-CCLP Ver. 3 closed-set response task would be greater than 80%.

Construct Validity

To assess whether SIP-CCLP Ver. 3 results discriminate between children with and without cleft palate, the following questions were posed with respect to construct validity: 1) How do results obtained from SIP-CCLP Ver. 3 compare for children with and without cleft palate? and 2) How do the results obtained from SIP-CCLP Ver. 3 compare across different age groups for a sample of children with typical speech development and no history of craniofacial abnormalities who vary in age? This second question was added to the original proposal since the anticipated number of children without cleft palate was recruited (n = 24) and were equally distributed across the four ages of interest. However, this was not the case for the children with cleft palate. Specifically, the following hypotheses were tested:

1. Children with cleft palate will have significantly lower intelligibility scores on the SIP-CCLP Ver. 3 open-set response task than age-similar peers without cleft palate.
2. Compared to children with typical speech development, on the closed-set response task, children with cleft palate will demonstrate:
    a. significantly lower percent target phonetic contrast items correct.
    b. significantly more cleft-related errors.
    c. significantly smaller percentage of obstruents correct.
    d. significantly greater percentage of correct items given a distorted rating.
3. Compared to older children, younger children will demonstrate:
    a. significantly lower intelligibility scores on the SIP-CCLP Ver. 3 open-set response task.
    b. significantly lower percent phonetic contrast items correct on the SIP-CCLP Ver. 3 closed-set response task.

Criterion Validity

To assess how well SIP-CCLP Ver. 3 predicts an outside criterion, in this case, spontaneous speech, the following questions were posed with respect to criterion validity: 1) What is the relationship between open-set intelligibility scores on SIP-CCLP Ver. 3

and the spontaneous speech sample? 2) What is the relationship between the speech sound error patterns on SIP-CCLP Ver. 3 and those in the spontaneous speech sample? Specifically, the following hypotheses were tested:

1. Intelligibility scores on the SIP-CCLP Ver. 3 open-set response task and those determined from the spontaneous speech sample will be moderately positively correlated.
2. Percentage consonants correct will be similar:
    a. between a phonetic analysis of SIP-CCLP Ver. 3 stimulus words and the spontaneous speech sample.
    b. between an error analysis of the SIP-CCLP Ver. 3 closed-set responses and the phonetic analysis of the spontaneous speech sample.

METHOD

Research Design

This study used several designs to test the hypotheses of interest. The research questions regarding construct validity involved comparisons between two groups of age-similar children with and without cleft palate and children without cleft palate of different ages. For the age-similar children with and without cleft palate, these comparisons were made using Student's t-tests for independent samples (one-tailed) for each dependent measure. For the children without cleft palate equally distributed across four age groups, these comparisons were made using one-way analyses of variance.

Assessment of criterion validity involved two designs. The relationship between intelligibility scores in the two sampling conditions (spontaneous and SIP-CCLP Ver. 3 words) was tested using linear regression analysis. The three sampling conditions (SIP-CCLP Ver. 3 phonetic transcription, SIP-CCLP Ver. 3 error analysis, spontaneous speech sample) were compared for the dependent variable, percent consonants correct, using a one-way repeated measures analysis of variance.

Child Participants

Recruitment

An estimate of the minimal sample size needed to ensure a power level of 0.8, using an alpha level of .05, was determined using a power analysis. The group mean and standard deviation values for the open-set intelligibility scores, closed-set intelligibility scores and percent obstruents incorrect for the eight children without cleft palate and four children with cleft palate described in Gotzke (2003) were used to estimate effect size following the procedures of Cohen, Welkowitz & Ewen (1981). This calculation is illustrated in Appendix G and resulted in effect sizes ranging from 1.0 to 1.1. Using the power tables found in Kirk (1968), with alpha = .05, power of .80, effect size between 1 and 1.5 and two groups (children with cleft palate and children with typical speech), the minimum number of subjects required per group was 17 for all three dependent variables. As data from five children with cleft palate and eight children with typical speech development had previously been collected by Feltz et al. (2002) and Gotzke (2003), the original recruitment goals were a minimum of an additional twelve children with cleft palate and nine children without cleft palate.

Parents were provided with an information letter describing the details of the project prior to the data collection appointment. Written consent was obtained from the parent and child on the testing date. Appendix I provides a sample of the information letter and consent form that was given to parents for the children recruited in the Edmonton area. A separate consent form was developed for the ethics application to recruit children from the Alberta Children's Hospital in Calgary. Parents were reimbursed for parking costs for the data collection session. Parents were also asked to provide information about their current occupation to allow determination of socioeconomic status. A socioeconomic index, namely a Blishen score, was assigned to each parent based on their occupations (Blishen, Carroll & Moore, 1987). In cases where both parents were employed, the Blishen scores were averaged to provide a single score for the child. A summary of the characteristics of children recruited for Feltz et al (2002), Gotzke (2003) and the current study are included in Table 5.

Inclusion Criteria

Children with Cleft Palate

Inclusion criteria for children with cleft palate included presence of a cleft palate (repaired or unrepaired; overt or submucous) as identified by the referring speech-language pathologist. Children with submucous cleft palate were included if the submucous cleft was identified by the referring speech-language pathologist and confirmed by a physician through the presence of a bifid uvula, notch at the end of the hard palate and a bluish line (i.e., zona pellucida) in the middle of the velum. Children who met the inclusion criteria for cleft palate also needed to have hearing within normal limits, as determined by an audiologist in sound field testing, and the absence of any concomitant physical and/or cognitive impairment as reported by the referral source and parents. Although ethical approval was obtained for inclusion of children with cleft palate and delayed expressive language abilities, for all children recruited, receptive and expressive language abilities were within one standard deviation of the mean for their chronological age based on the child's score from the auditory comprehension and vocal expression subtests of the Preschool Language Scale, Fourth Edition (PLS-4) (Zimmerman, Steiner & Pond, 2002).

Data were collected from nine children with cleft palate ranging in age from 37 to 72 months who were recruited from clientele at the Alberta Children's Hospital Cleft Palate Clinic. Two of the children (CP04 and CP12) had diagnosed syndromes and so did not meet criteria for inclusion in the group of children with cleft palate. A third child, who was 37 months of age, started but then lost interest in the task and did not complete the SIP-CCLP recording. A tenth child with cleft palate, aged 39 months, was recruited from the University of Alberta Cleft Palate Clinic. Of the 10 children recruited, seven met all inclusion criteria and had complete data for analysis. These seven children's data, combined with those five children with cleft palate from the two previous studies (Feltz et al., 2002; Gotzke, 2003), who also met all criteria and had complete data recordings, resulted in a group of 12 children. In this group, five children had a unilateral cleft lip and palate, three children had a cleft of the soft palate only, one child had a bilateral cleft lip and palate, one child had an incomplete cleft palate, one child had a cleft of the hard and soft palate and one child had a submucous cleft. All clefts had been repaired surgically.

Table 5.

Subject Pool Characteristics for Children Recruited for Feltz et al. (2002), Gotzke (2003) and the Current Study.

	Children with Cleft Palate			Children without Cleft Palate		
Age Range in Months	Subject Number	Age in Months	Gender	Subject Number	Age in Months	Gender
36 – 47	CP01	39	M	TS01[1]	37	F
	CP02	39	F	TS02	39	F
	CP03[2]	41	F	TS03	39	F
	CP04	***41***	***M***	TS04	40	F
	CP05[2]	45	F	**TS05[2]**	**40**	**M**
	CP06[2]	47	M	TS06[2]	43	F
	CP07	47	M	TS07	44	M
				TS08	46	F
48 - 59	CP08	48	F	TS09	51	F
	CP09	50	F	**TS10**	**51**	**M**
	CP10	56	F	TS11	53	M
	CP11[1]	59	M	TS12	56	F
				TS13[2]	57	M
				TS14[2]	57	M
				TS15	59	F
60 - 71	***CP12***	***60***	***F***	**TS16**	**60**	**M**
	CP13	64	F	TS17[2]	60	M
				TS18[1]	61	F
				TS19	61	F
				TS20[1]	65	M
				TS21[2]	**67**	**F**
				TS22	68	F
				TS23	71	F
72 - 84	CP14[1]	79	M	**TS24[1]**	**73**	**M**
				TS25	75	M
				TS26	75	M
				TS27	75	M
				TS28[1]	**78**	**M**
				TS29	79	F
				TS30[1]	82	M
				TS31[2]	84	F

Note. [1]Subjects from Feltz et al (2002) [2]Data from Gotzke (2003). Bolded and italicized type indicates those children with cleft palate with a diagnosed syndrome. Bolded type indicates those children without cleft palate identified as having an articulation disorder.

Children without Cleft Palate

For inclusion in the study, each child without cleft palate was required to pass a pure tone hearing screening (ASHA, 1985 guidelines) and an oral mechanism screening using the Dworkin-Culatta Oral Mechanism Exam (D-COME-T) (Dworkin & Culatta, 1996) and a speech and language screening. The Fluharty Preschool Speech and Language Screening Test (Fluharty -2) (Fluharty, 2001) was used to screen articulation, and receptive and expressive language abilities. To pass the screening, children had to score at or above the 16th percentile. Screening procedures for the children without cleft palate took approximately 15 minutes to complete.

As recruitment of children with cleft palate was challenging, the recruitment goal for children without cleft palate of different ages was increased to 24 children, evenly distributed across four age groups: 36 - 47 months, 48 - 59 months, 60 - 71 months and 72 – 84 months to test an additional hypothesis about the effect of age on SIP-CCLP scores. As Feltz et al. (2002) and Gotzke (2003) collected data for two or three children in each of the above age groups, the following recruitment goals were set: 4 children (36 - 47 months), 4 children (48 - 59 months), 3 children (60 – 71 months), and 4 children (72 – 84 months), for a total of 15 children. This recruitment goal was met. However, two children, TS10 and TS16, did not meet the screening criteria, therefore, two additional children were recruited via information posters and leaflets distributed in the community, increasing the total number of children recruited to 17. As all children in the 36 – 47 month group who met the inclusion criteria were female, an additional male was recruited for this group, increasing the total number of children recruited to 18. As shown in Table 5, four children (TS05, TS21, TS24 and TS28) from Feltz et al. (2002)

and Gotzke (2003) and TS10 and TS16 did not pass the articulation screening. Data from these children were excluded from the data analysis. In total, data from 25 children who met the inclusion criteria for the children without cleft palate were available for analysis.

Speech Sampling Conditions

Spontaneous Sample

Data collection sessions were conducted in either a sound booth or a quiet room, depending on the site. Using the procedure outlined by Shriberg (1986), a spontaneous speech sample was collected from each child using a parallel play and an interactive play scenario. Speech samples were approximately fifteen minutes in length. Play dough was used to stimulate language production. In addition, the examiner introduced at least one topic that was unrelated to the play materials during the interaction to further stimulate production. The examiner used both questions and comments to prompt verbalizations about the stimulus materials. The spontaneous speech sample was recorded directly to the computer hard drive (personal computer using Windows XP operating system with a SoundBlaster sound card) as a digital audio file. The recording was captured using a Shure WH20 unidirectional dynamic headset microphone, an Audio Buddy Dual Mic Preamplifier and custom software (TOCS+ Recorder) with a sampling rate of 48 KHz and a quantization size of 16 bits. The spontaneous speech sample was also recorded using a Panasonic AG-196 video camera and a Sony Electret lapel microphone for back-up purposes and to provide additional contextual information in glossing the child's utterances. One child with cleft palate (CP02) refused to wear the headset microphone for the spontaneous sample, so the audio track of the sample recorded using the video camera and lapel microphone was used for analysis.

SIP-CCLP Ver. 3

SIP-CCLP Ver. 3 was administered to each child following elicitation of the spontaneous speech sample. Prior to the testing session, order of item presentation was randomized using SIP-CCLP Ver. 3 software to create a unique test order for each child. Picture stimuli were presented on a 15-inch screen with 800 X 600 resolution. The child was instructed to repeat the name of the picture displayed after the examiner modeled the word. Four practice words preceded the presentation of the actual stimulus words. If the child exhibited any confusion with the task, additional instructions were provided. Once the child demonstrated comprehension of the task, presentation of the actual stimulus words began with the examiner eliciting the target words. Short breaks were provided in the form of musical computer animations that appeared after every twenty stimulus words. All 124 stimulus words were presented with the examiner modeling the word(s) for the child to repeat. If the examiner was unsure about the recording quality of any item or had any concerns about background noise or examiner voiceover, a second imitation was elicited. To further support the children's understanding of the imitation task, visual and auditory cues were used to indicate to the child when it was his turn to repeat the word. A short beep and the appearance of a puppet on the top left corner of the screen were used to indicate when it was the child's turn. Verbal reminders to "wait for the beep before speaking" were also given.

Administration of SIP-CCLP Ver. 3 to the child participants took between 10 and 20 minutes. The child's attention to the task and the number of times that items had to be repeated to ensure a clean recording affected the amount of time required to complete the task. The examiner provided short breaks as needed to maintain the child's engagement with the task.

Word productions were recorded directly to the computer as digital audio files using the same instrumental set-up as for the spontaneous speech sample (Shure WH20 unidirectional dynamic headset microphone and an Audio Buddy Dual Mic Preamplifier connected to the microphone input of the sound card). Administration of SIP-CCLP Ver. 3 was also recorded using a Panasonic AG-196 video camera connected to a Sony Electret lapel microphone, for back-up purposes. As for the spontaneous speech sample, when possible, administration of the SIP-CCLP Ver. 3 occurred in a sound booth. Alternately, a quiet room was used.

Assessment of Validity and Reliability

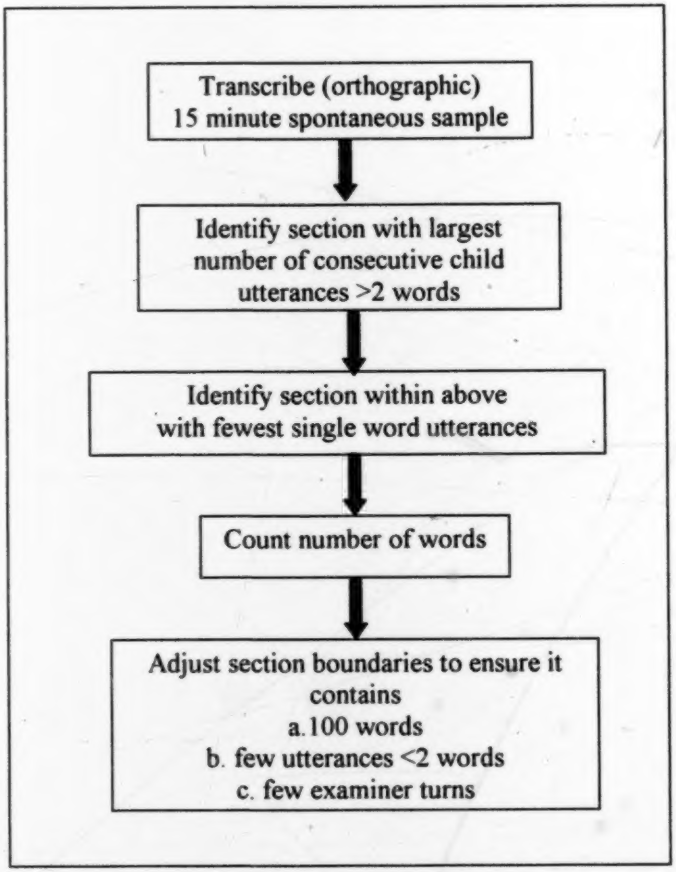
Preparation of Recordings for Listener Judges

Spontaneous Sample

The examiner orthographically transcribed each utterance in each child's 15 minute spontaneous speech sample using Adobe Audition (Adobe Systems Incorporated, 2004) to playback the digital recording. When a word or phrase could not be understood, the unintelligible utterance was played back multiple times. Words or phrases that could not be glossed, despite repeated playback, were indicated in the transcript by "X". Utterance boundaries were determined using the conventions of Shriberg (1986).

In the orthographic transcription, the examiner found a section with multiple child utterances longer than 2 words and few examiner turns. From this section in the transcript, the examiner selected a subsection containing 100 consecutive words and few utterance boundaries as per Weiss (1980). This 100-word sample was used for the listening task. A flowchart that illustrates the procedures for selecting the 100-word sample is shown in Figure 2.

Figure 2. Flowchart for selection of 100-word spontaneous sample.


Transcribe (orthographic)
15 minute spontaneous sample
Identify section with largest number of consecutive child utterances >2 words
Identify section within above with fewest single word utterances
Count number of words
Adjust section boundaries to ensure it contains
a. 100 words
b. few utterances <2 words
c. few examiner turns

As CP02 refused to wear the headmount microphone for the spontaneous sample, the sample was transcribed (orthographic) from the video recording. Selection of the 100-word sample followed the protocol described above. For CP02, each utterance in the100-word sample was digitally recorded from the video recording using CSpeech 4.0

(Milenkovic & Read, 1994) with a 16 bit quantization size and a 22 kHz sampling rate. These digital files were converted from CSpeech files to audio wave files for playback.

During playback, each of the child's utterances was presented as a separate audio-file in the order of occurrence in the transcript. The orthographic gloss was used as the key against which the listener judge's responses were compared for scoring.

SIP-CCLP Ver. 3

The examiner listened to each child's SIP-CCLP Ver. 3 word recordings using Adobe Audition (Adobe Systems Incorporated, 2004) to playback the digital .wav files. Multiple productions of the target word, environmental noise and examiner and child comments were edited from the recordings using Adobe Audition. When there were multiple productions of the target word, the first production without examiner voiceover or environmental noise interference was saved as the .wav file for playback to listener judges.

Listener Judges

Recruitment

Listener judges were recruited from students in the Faculty of Rehabilitation Medicine at the University of Alberta. Each listener had Canadian English as their first language and normal hearing as determined by a hearing screening performed by the researcher according to ASHA (1985) guidelines. A total of 78 listener judges were recruited. An honorarium of ten dollars was given to each listener for their time and participation. Listener judges were provided with an information letter at the beginning of the listening session. Written consent was also obtained. A sample of the information letter and the consent form are provided in Appendix J.

Two listeners from the Department of Speech-Language Pathology (one first-year and one second-year) and one student from either Occupational Therapy or Physical Therapy were randomly assigned to judge each child subject's SIP-CCLP Ver. 3 recordings and a second set of three listeners (two from speech-language pathology; one from occupational or physical therapy) were randomly assigned to judge each child's spontaneous speech sample. Listeners did not hear the same child's SIP-CCLP Ver. 3 and spontaneous sample recordings. In instances where listener judges participated in more than one listening session, a minimum of three weeks elapsed between listening sessions.

Instructions

All listening sessions took place in a Madsen OB822 sound booth. During the listening task, the computer hard drive (personal computer with Windows XP operating system with a SoundBlaster sound card) was set up outside the sound booth to improve the signal-to-noise ratio. Speech samples were presented through a Technics Stereo Integrated Amplifier (model SU-V460) connected to ElectroVoice S-40 compact monitor speakers located in the sound booth. The playback volume of SIP-CCLP Ver. 3 stimulus words and the spontaneous speech sample was standardized prior to the listening sessions to ensure all listener judges heard the productions at approximately the same volume. Play back volume was set such that the practice items were heard at 55 – 65 dBA. Listener judges were asked about the comfort level of the playback volume after the practice items for all three tasks (open-set, closed-set, and spontaneous speech sample) and adjustments made if necessary.

SIP-CCLP Ver. 3 word identification – open-set response task.

At the beginning of the listening session for SIP-CCLP Ver. 3, listener judges were asked to read over a list of words presented on the screen two times to familiarize them with the kinds of words to be heard and to reduce the effect of differences in listener familiarity with the test words (Yorkston & Beukelman, 1981). Appendix K contains the list of priming words read by the listener judges. After the word familiarization task, the children's recordings were presented. Each listener judge was instructed to type the word or words heard into the computer. If the words were not clear, the listener judge was instructed to try and make her best guess as to the words the child said. If the listener judge had no idea what was said, she was instructed to type in whatever sound or sounds were recognized. Instructions were also presented on the computer screen. The researcher answered any questions once the instructions had been read and, again, following the practice items. Any volume adjustments necessary to ensure a comfortable playback level were also made after the four practice items.

For the open-set response task, the SIP-CCLP Ver. 3 software randomly selected twelve words from the 124 stimulus words for a second presentation. Repeated stimulus words were ordered randomly throughout the set of 124 stimulus words. The entire open-set response task consisted of four practice words, the 124 stimulus words and twelve repeated items, bringing the total number of response judgments to 140 items. For each listener, the open-set response task was administered in its entirety with no repetitions or breaks. The open-set response task took approximately 15 minutes to complete. The listener judge was allowed a break following completion of the open-set response task.

SIP-CCLP Ver. 3 phonetic contrast identification – closed-set response task.

Following the break, the listener judge was presented with instructions for the closed-set response task. Listener judges were told that they will see two words with a sound or sounds underlined. They were instructed to focus on what was heard in the underlined positions in the word as they heard the words spoken by the child. If what was heard corresponded to one of the two underlined choices present, the listener judge was instructed to select it. If a different sound(s) was heard in the underlined position, listener judges were instructed to select the blank button and type in what was heard. If the sound(s) heard in the underlined position could not be identified, listener judges were instructed to select the "can't identify" button.

Listener judges were also instructed to rate what was heard in the underlined position as "clear" or "distorted" if one of the first three buttons (a member of the minimal contrast pair or the blank) were selected. A distortion rating was not requested if the listener judge chose the "can't identify" response. The investigator answered any questions that the listener judge had after the instructions had been given and the four practice items were completed.

For the closed-set response task, the SIP-CCLP Ver. 3 software randomly generated the order of presentation for the contrast items. The child's word productions that tapped more than one error pattern were presented more than once. For example, the listener judge heard the child's production of the stimulus word "rail" four times. Each repeated production was paired with different response choices on the closed-set response task computer screen. For example, each of the following four phonetic contrast pairs were provided as possible choices once in the closed-set response task when the child's production of "rail" was played: "rail/tail", "rail/pail", "rail/jail" and "rail/trail." The SIP-CCLP Ver. 3 software randomly selected 20 target phonetic contrast items for a second presentation. These reliability items were presented randomly throughout the response task. The closed-set response task consisted of four practice items, 194 target phonetic contrast items and 20 repeated contrast items for a total of 218 items and took approximately 18 minutes to complete. The listener judge was allowed a short break following completion of the closed-set response task.

Spontaneous sample.

Following the second break, listener judges were informed that they would now hear a speech sample consisting of a number of utterances varying in word length. They were also told that this sample would not be from the same child as in the previous two listening tasks. Listener judges were instructed to write down what they heard. Listener judges heard each utterance twice if desired. The SIP-CCLP Ver. 3 practice words were heard at the beginning of the task to prime the listener judges to the new child's voice. Adjustments to the playback volume were made if necessary. This task took approximately 7 minutes to complete. The listener judge was then thanked for his/her participation and provided with an honorarium.

Measurement of Dependent Variables

A brief description of the dependent variables in this study is included for reference in Appendix K.

Spontaneous Sample

Intelligibility analysis.

The number of words identified correctly by each listener judge was determined from a comparison of the listener's written responses with the orthographic gloss (key) prepared by the investigator. This number was converted to a percentage (number identified correctly/ total number of words in the sample X 100). The mean percentage of the three listener judges served as the subject's intelligibility score.

Phonetic analysis.

Phonetic transcription of the 100-word sample was conducted by the researcher to determine percent consonants correct. In transcribing the speech sample, the researcher followed guidelines outlined by Shriberg (1986). Broad phonetic transcription, as well as narrow transcription using diacritic symbols for the following distortions: dentalized, palatalized, lateralized, nasalization and nasal emission, was used.

To determine interjudge agreement for the phonetic transcription, each 100-word sample was transcribed independently by a second trained transcriber. Listener agreement on the phonetic transcription was based on point-by-point comparison of the consonants only for the two transcripts. Instances where both transcribers indicated that a consonant was distorted but did not agree on the diacritic were counted as an agreement. Inter-transcriber agreement for the phonetic transcription (broad and subset of narrow diacritics for distortions) of the consonants in the 100-word spontaneous sample for the 12 children with cleft palate was 77.0% (*SD* = 5.6) and for the 25 children with typical speech was 88.7% (*SD* = 5.1). Overall, inter-transcriber agreement was 85.0% (*SD* = 7.6). Shriberg and Lof (1991) reported average interjudge agreement for broad and narrow transcription of 89% and 74% respectively for consonants in a continuous speech sample for a group of speakers with and without speech disorders. For the diacritics used in this study, Shriberg and Lof (1991) reported average interjudge agreement of 67.1%. In the current study, less than 74% inter-transcriber agreement was obtained for 2 of the 12 children with cleft palate (CP05 = 65.8%; CP11 = 71.6%).

SIP-CCLP Ver. 3

Intelligibility analysis.

For the open-set response task, the intelligibility score for each subject was calculated in the same way as for the spontaneous sample. A mean of the percentage words correct for the three listener judges was calculated and served as the child's intelligibility score.

Intrajudge agreement for the SIP-CCLP Ver. 3 open-set response task was determined using the judges' responses to the twelve repeated items. Each listener judge's responses to the first and second presentation were compared for agreement. The number of agreements was determined for each listener judge. The mean number of agreements for the three listener judges was calculated for each subject.

For the closed-set response task, the printout of results for each listener judge was examined to determine the number of items for which an error code (ECode) of "1" (correct/clear) or "2" (correct/distorted) was given. The percentage of words identified correctly (i.e., "1"s and "2"s) was determined for each listener judge. A mean of these percentages for the three listener judges served as the subject's intelligibility score for the closed-set response task. Figure 3 shows an example of the results printout obtained for the SIP-CCLP Ver. 3 closed-set response task.

Figure 3. Example of results for the SIP-CCLP Ver. 3 closed-set response task.

	Code	Actual Word	Listener Choice	RT[1]	ECode[2]
1.	MPC.SLi. I. 7	toe	toe	2.656	2
2.	SDU.LA.M.7	nut fell	“nut bell”	3.953	5
3.	VEC.S.UVo.F.9	lock	lock	2.906	1
4.	MPC.SN.M.32	muddy	“muggy’	2.891	6
5.	MPU.GIF.I.13	well	well	2.437	1
6.	SDC.AP.M.3	nut sell	nut shell	1.844	4
7.	PPC.AB.M.7	muddy	mug E	1.813	3
8.	VED.S.UVo.I.3	tear	CI	5.391	7
.					
.					
.					
194.	PPC.AB.I.5	tail	tail	1.687	2

Note. [1]RT= reaction time in seconds. [2]ECode = Error Code for listener's response. Information about closed-set codes can be found in Appendix B.

Intrajudge agreement for the judgments of “correct/incorrect” for the SIP-CCLP Ver. 3 closed-set response task was determined using the judges’ responses to twenty repeated phonetic contrast items. As this feature was added to SIP-CCLP Ver. 3, determination of intrajudge agreement was possible only for those children whose data were collected using this version: 7 children with cleft palate and 16 children without cleft palate. For these 23 children, each of the listener judge’s responses to the first and second presentation were compared to see if they chose or typed in the same response for each of 20 agreement items. If a listener judge rated a production as correct/clear (i.e., “1”) in the first presentation and correct/distorted (i.e., “2”) in the second presentation or vice versa, the item was considered an identification agreement. Similarly, if a rating of incorrect/clear (i.e., “3”) was given in the first presentation and incorrect/distorted (i.e., “4”) was given in the second presentation (or vice versa), the item was considered a misidentification agreement. For the typed-in responses (i.e. “5” and “6”), listener judges had to type in the exact same word on both presentations of the phonetic contrast to be considered an agreement. The number of agreements out of 20 was determined for each listener judge. The mean number of agreements for the three listener judges was calculated for each subject.

Phonetic analysis.

Phonetic transcription of the 124 SIP-CCLP Ver. 3 words was conducted and analyzed using P.E.P.P.E.R. (Shriberg, 1986) to determine percent consonants correct. In transcribing the child's SIP-CCLP Ver. 3 recordings, the researcher followed the same guidelines as described for the phonetic transcription of the spontaneous speech sample including a combination of broad and narrow transcription for the subset of diacritics for distortions (i.e., dentalized, palatalized, lateralized, nasalization and nasal emission).

To determine interjudge agreement for the phonetic transcription, a random sample of 20% of the words from each child's SIP-CCLP Ver. 3 recording was transcribed independently by a second trained transcriber. Listener agreement on the phonetic transcription was based on point-by-point comparison of the consonants only of the two transcripts. Instances where both transcribers indicated that a consonant was distorted but did not agree on the diacritic were counted as an agreement. Inter-transcriber agreement for the broad and selected narrow phonetic transcription of the consonants in the phonetic transcription of 20% of the SIP-CCLP Ver. 3 words for the 12 children with cleft palate was 81.3% (*SD* = 8.7) and for the 25 children with typical speech was 91.1% (*SD* = 5.0). Overall, inter-transcriber agreement was 87.9% (*SD* = 7.8). Shriberg and Lof (1991) reported average interjudge agreement for broad and narrow transcription of 89% and 61% respectively on consonants in an articulation test for a group of speakers with and without speech disorders. For the diacritics used in this study, Shriberg and Lof (1991) reported average interjudge agreement of 67.1%. Less than 61% inter-transcriber agreement was obtained for 1 of the 12 children with cleft palate (CP10 = 60.4%).

Closed-set: Phonetic contrast error analysis.

Error pattern profiles were generated for each subject based on listener responses on the closed-set response task. SIP-CCLP Ver. 3 closed-set target phonetic contrast items were examined to determine those items for which a minimum of two of the three listeners chose the "foil" (i.e., "3" or "4") or typed in the same response in the "blank" (i.e., "5" or "6"). These items were considered as errors and were used to create error profiles. For example in Figure 4, the target phonetic contrast items VEU.S.VoU.I.2 and PPC.AV.I.1 would be considered as errors as a minimum of 2 of the 3 listener judges chose the same foil for the contrast, "pail" and "gown" respectively. Similarly, VEU.S.VoU.I.3 would be considered as an error as 2 of the 3 listener judges typed the same response in the blank, "gear". Listener judges' responses on the SIP-CCLP Ver. 3 closed-set were also examined to determine those items in which a minimum of 2 of the 3 listeners chose something other than the correct response but did not agree on their response. These items were also considered errors. For example in Figure 4, for the target phonetic contrast CEU.9, one judge chose the foil, one judge chose the target and one judge typed in a response that was different from both the foil and the target. Those items in which a minimum of 2 of 3 listeners chose "can't identify" were also considered as errors. Figure 4 illustrates an example of this kind of error for the target phonetic contrast, CED.8. The total number of errors was determined for each child participant and converted to a percentage of the total number of items in the closed-set response task (n = 194 for SIP-CCLP Ver. 3). For each child, incorrect items were analyzed to determine the following: 1) number items in which 2 of 3 listeners chose the foil (i.e., "3"s or "4"s), 2) number of items in which 2 of 3 listeners typed the same response in the "blank" (i.e., "5"s or "6"s) 3) number of items in which 2 of 3 listeners chose "can't identify" (i.e., "7"s) and 4) number of items in which no consensus was reached.

Figure 4. Example of collated listener responses on the SIP-CCLP Ver. 3 closed-set response task.

Target Phonetic Contrast	Word Played	ECode			Listener Choice		
		Listener Judge 1	Listener Judge 2	Listener Judge 3	Listener Judge 1	Listener Judge 2	Listener Judge 3
GEU.?V.7	ape	4	1	1	cape	ape	ape
GEC.B?.I.1	B	1	1	1	B	B	B
VEU.S.VoU.I.2	bale	3	4	3	pail	pail	pail
VEU.S.VoU.1.3	deer	5	5	2	"gear"	"gear"	deer
PPC.AV.1.1	down	4	1	4	gown	down	gown
MPU.LiF.I.16	lace	2	1	6	lace	lace	"nace"
MPU.LiF.I.21	rip	5	5	1	"whip"	"whip"	rip
CED.8	string	7	6	7	"CI"	"fring"	"CI"
CEU.9	tick	4	1	5	stick	tick	"kick"
CED.4	trail	5	6	6	"fail"	"fair"	"fair"

Interjudge reliability on items in the closed-set response task was determined by identifying the number of items for which all 3 listeners identified the target correctly (i.e., correct/clear ("1") and correct/distorted ("2")) or all 3 listeners selected an "incorrect" response (i.e., foil/clear ("3"), foil/distorted ("4"), blank/clear ("5"), blank/distorted ("6") and can't identify ("7")). This number of agreements was the divided by the total number of judgments made and converted to a percentage for each child.

Errors were then classified according to three categories described for SIP-CCLP Ver. 3 (i.e., cleft-related, developmental, unknown). For the errors in which 2 of 3 listeners chose the foil, the error was classified according to the code for that target phonetic contrast. For example, in Figure 4, the target phonetic contrast VEU.S.VoU.I.2 was identified as an error as all 3 listeners chose the foil "pail". As this was the provided foil, the error was identified as an "Unexpected voicing error", based on the first three characters in the code for the contrast. For errors in which a minimum of 2 of the 3 listener judges typed in the same response, the error pattern identified was compared with the target phonetic contrast items tested in SIP-CCLP Ver. 3 (see Appendix B). If a similar target phonetic contrast was found, the error was classified according to the first three characters in the code for the contrast. For example, in Figure 4, the target phonetic contrast VEU.S.VoU.I.3 was identified as an error when 2 of the 3 listeners typed in the response "gear". As this is not a voicing error, the list of target phonetic contrast items in Appendix B was examined to determine if there were any contrast items in which alveolar (/d/) was contrasted with a velar (/g/). As cleft-related place preference errors sample similarly contrastive pairs, the error was classified as a "Cleft-related place preference error".

If a similar target phonetic contrast was not found, the contrast pair was examined to see it represented a cleft-related or developmental error pattern not tested in SIP-CCLP Ver. 3 and categorized as "other". A subcategory of "other" was added to include those errors identified by listeners that were cleft-related or developmental but did not fit into one of the predetermined subcategories. For example, in Figure 4, the target phonetic contrast MPU.LiF.I.21 was identified as an error when 2 of the 3 listeners typed in the response "whip". As no phonetic contrast items in SIP-CCLP Ver. 3 test developmental gliding errors as represented by the pair "rip – whip", this error was classified as a "Developmental other error".

If a similar target phonetic contrast was not found and the pattern did not represent a cleft-related or developmental error not tested in SIP-CCLP Ver. 3, the error was called "unclassified". This category includes those errors not identifiable as belonging to one of the predetermined categories in SIP-CCLP Ver. 3. For example, the contrast pair "trail - fair" identified by listeners for target contrast CED.4 in Figure 4 is not identifiable as either cleft-related or developmental and would then be considered "unclassified".

Number of cleft-related, developmental, unknown and "unclassified" errors were calculated for each subject. Cleft-related and developmental errors were further examined to determine the number of sibilant errors, voicing errors, glottal errors (cleft-related only), manner preference errors, place preference errors, cluster errors (developmental only) and "other" in each category for each subject.

Closed-set: Sound analysis of error patterns.

For each child, the target sound in each error was identified as being either a sonorant or obstruent. The total number of obstruents identified incorrectly was determined and subtracted from the total number of obstruents targeted in SIP-CCLP (Ver. 3, n = 141). This value was then divided by the total number of obstruents targeted in SIP-CCLP Ver. 3 and converted to a percent. This value served as the percent obstruents correct for the closed-set response task for each subject. Percent sonorants correct were also determined following the above protocol (SIP-CCLP Ver. 3, number of sonorants = 44). Each obstruent and sonorant in error was also examined to determine its manner, place and voicing characteristics. Following the same protocol as for calculating percent obstruents correct, percent correct for each manner class (i.e., stops, fricatives, affricates, nasals, liquids, glides), places of articulation (i.e., bilabial, labiodental, interdental, alveolar, palatoalveolar, velar, glottal) and voicing (i.e., voiced, voiceless) were also determined.

Closed-set: Distortion analysis.

Percentage of correct/distorted and incorrect/distorted items were calculated by determining the mean percent correct/distorted (i.e., "2"s) and incorrect distorted (i.e., "4"s and "6"s) responses for the three listeners for each child. Interjudge agreement for the distortion ratings was evaluated by determining the number of items for which all three listeners gave the same rating of clear (i.e., "1", "3" or "5") or distorted (i.e., "2", "4" or "6") and then dividing this number by the total number of items judged and then converting the result to a percentage.

Intrajudge agreement for the ratings of "distorted/clear" for the SIP-CCLP Ver. 3 closed-set response task was determined using the judges' responses to twenty repeated phonetic contrast items. As this feature was added to SIP-CCLP Ver. 3, determination of intrajudge agreement was possible only for those children whose data were collected using this version: 7 children with cleft palate and 16 children without cleft palate. For these 23 children, each of the listener judge's responses to the first and second presentation were compared to see if they chose the same rating of "clear" or "distorted" for the 20 agreement items. If a listener judge rated a production as correct/clear (i.e., "1") in the first presentation and incorrect/clear (i.e., "3") in the second presentation or vice versa, the item was considered a distortion agreement. If a rating of "blank"/distorted (i.e., "6") was given in the first presentation and incorrect/clear (i.e., "3") was given in the second presentation (or vice versa), the item was not considered a rating agreement. The number of agreements out of 20 was determined for each listener judge. The mean number of agreements for the three listener judges was calculated for each subject.

Analysis

Data from the two children with cleft palate described by Gotzke and Hodge (2003) and three children with cleft palate described by Gotzke (2003) were combined with data collected using SIP-CCLP Ver. 3 for 7 children with cleft palate to evaluate validity and reliability of the SIP-CCLP measure. Table 6 provides information about the children with cleft palate.

Data from 25 children with typical speech development were also used to evaluate the validity and reliability of the SIP-CCLP measure. This data set includes four children with typical speech development reported by Feltz et al. (2002) and five children with typical speech development reported by Gotzke (2003). From this pool of 25 children, 12 children similar in age to the children with cleft palate were selected. Table 6 also contains information about these 12 children.

Blishen scores (Blishen, Carroll & Moore, 1987) were given to each child based on parental occupation. Scores were not available for four of the children with cleft palate and one child without cleft palate. The mean Blishen score for the groups of children with and without cleft palate were 44.3 (*SD* = 11.7) and 48.3 (*SD* = 10.4). Blishen scores between the two groups of children were similar (t = -0.78, p = 0.442).

To compare scores on SIP-CCLP for younger and older children with typical speech age, six children in each of the age intervals (i.e., 36 – 47 months; 48 – 59 months, 60 – 71 months, 72 - 84 months) were selected from the pool of children with typical speech. As data from 25 children with typical speech was collected, data from one child (TS2) was excluded from this analysis as this child was a twin to TS3. Table 7 contains information about each child in this group.

Table 6.

Subject Information for Children with Cleft Palate and their Age-Similar Peers without Cleft Palate

	Children with Cleft Palate					Children without Cleft Palate and with Typical Speech Development			
	ID	Age[6]	Gender	Cleft Type	SES (Blishen Score)	ID	Age[6]	Gender	SES (Blishen Score)
3,0 - 3,11	CP01	39	M	Cleft of Soft Palate Only	48.01	TS01[1]	37	F	N/A[5]
	CP02	39	F	Submucous	63.48	TS02	39	F	56.83
	CP03[2]	41	F	Cleft of Soft Palate Only	43.29	TS03	39	F	56.83
	CP05[2]	45	F	BCLP[3]	N/A	TS06[2]	43	F	59.44
	CP06[2]	47	M	UCLP[4]	N/A	TS07	44	M	50.66
	CP07	47	M	Incomplete Cleft Palate	44.41	TS08	46	F	37.36
4.0 – 4,11	CP08	48	F	Cleft of Soft Palate Only	31.93	TS09	51	F	34.86
	CP09	50	F	Cleft of Hard and Soft Palate	56.83	TS11	53	M	41.42
	CP10	56	F	UCLP	34.85	TS12	56	F	44.92
	CP11[1]	59	M	UCLP	N/A	TS13[2]	57	M	41.69
5,0 - 5,11	CP12	64	F	UCLP	31.22	TS19	61	F	39.99
6,0 – 6,11	CP14[1]	79	M	UCLP	N/A	TS27	75	M	66.85
Group *M* (*SD*)		51.2 (11.7)	7F: 5M		44.3 (11.7)	Group *M* (*SD*)	50.1 (11.1)	8F: 4M	48.3 (10.4)

Note. [1]Subjects from Feltz et al. (2002) [2]Data from Gotzke (2003) [3]Bilateral Cleft Lip and Palate [4]Unilateral Cleft Lip and Palate [5]Information on parental occupation was not collected, therefore, Blishen scores were not available for these children. [6]Age in months.

Table 7.

Subject Information for Children without Cleft Palate

Age Range in Months	Subject Number	Age in Months	Gender
36 – 47	TS01[1]	37	F
	TS03	39	F
	TS04	40	F
	TS06[2]	43	F
M = 41.5	TS07	44	M
SD = 3.4	TS08	46	F
48 – 59	TS09	51	F
	TS11	53	M
	TS12	56	F
	TS13[2]	57	M
M = 55.5	TS14[2]	57	M
SD = 2.9	TS15	59	F
60 – 71	TS17[2]	60	M
	TS18[1]	61	F
	TS19	61	F
	TS20[1]	65	M
M = 64.3	TS22	68	F
SD = 4.5	TS23	71	F
72 – 84	TS25	75	M
	TS26	75	M
	TS27	75	M
	TS29	79	F
M = 78.3	TS30[1]	82	M
SD = 4.0	TS31[2]	84	F

Note. [1]Subjects from Feltz et al (2002) [2]Data from Gotzke (2003).

For the children from Feltz et al. (2002), recordings collected for the SIP-CCLP Ver. 1 and spontaneous sample were reanalyzed for this study. SIP-CCLP Ver. 1 recordings were played to listeners using the computerized SIP-CCLP Ver. 3 and analyzed accordingly. As nine words were not elicited from these children, intelligibility scores for the SIP-CCLP Ver. 3 open-set response task were based on 115 words. For the SIP-CCLP Ver. 3 closed-set response task, analysis was based on 182 phonetic contrast items. The spontaneous samples collected for these children were prepared and analyzed using the methods described for the other subjects in this study. As data from children from Feltz et al (2002) and Gotzke (2003) are included in the following analysis, results will be discussed relative to the SIP-CCLP measure, as opposed to Version 3 which was used in this study.

Reliability

SIP-CCLP open-set response task.

Interjudge reliability was estimated via an intraclass correlation coefficient (Shrout & Fleiss, 1979) for the three listeners' scores. A strong positive ICC was expected. Intrajudge agreement was calculated by determining the number of agreements on the 12 repeated items.

SIP-CCLP closed-set response task.

Interjudge reliability was estimated via an intraclass correlation coefficient (Shrout & Fleiss, 1979) for the three listeners' scores. A strong positive ICC was expected. Intrajudge agreement for the judgments of "correct/incorrect" and ratings of "clear/distorted" was calculated by determining the number of agreements on the 20 repeated items. For the closed-set response task, interjudge reliability across the three listeners for judgments of "correct/incorrect" and ratings of "clear/distorted" were expected to be greater than 80%.

Construct Validity

Children with and without cleft palate.

Children with cleft palate were expected to have lower intelligibility scores on the open-set response task. On the closed-set response task, these children were expected to have a higher percentage of errors, lower percentage of obstruents correct, and a greater percentage of correct items given a distorted rating. Each of these hypotheses was tested using a Student's t-test for independent samples (one-tailed). Children with cleft palate were also expected to have a greater number of cleft-related errors. This hypothesis was tested using a Mann-Whitney U test. Effect size was calculated using the procedure illustrated in Appendix G.

Children without cleft palate of different ages.

Younger children with typical speech development were expected to have lower intelligibility scores on the open-set response task. On the closed-set response task, these children were expected to have a lower percent phonetic contrast items correct than older children with typical speech development. Each of these hypotheses was tested using an one-way ANOVA.

Criterion Validity

To examine the relationship between intelligibility scores on the open-set response task and from the spontaneous speech sample, a Pearson Product-Moment correlation coefficient was calculated. It was expected that these variables would be moderately positively correlated.

Percent consonants correct scores were expected to be similar for the phonetic transcription of the SIP-CCLP stimulus words, the phonetic transcription of the 100-word spontaneous speech sample and the error analysis of the SIP-CCLP closed-set responses. This hypothesis was tested using a one-way repeated measures ANOVA.

RESULTS

Reliability and Construct Validity

Comparison of Children with and without Cleft Palate

Intelligibility Analysis

SIP-CCLP open-set response task.

On the SIP-CCLP open-set response task, intelligibility scores for the 12 children with cleft palate (*M* = 53.3%, *SD* = 15.9) were significantly lower than the intelligibility scores for the 12 age-similar children without cleft palate (*M* = 76.7%, *SD* = 10.1; t = -4.31, p < .000) with an effect size of 1.25. These results are shown in Figure 5. Intelligibility scores ranged from 26.9% to 76.8% for the children with cleft palate and from 55.9% to 88.9% for the children without cleft palate.

Intrajudge agreement, based on the 12 repeated items, was 84.3% (10.1/12, *SD* = 1.3) for the 72 listeners. When examined by group, mean intrajudge agreement was 76.5% (9.2/12, *SD* = 1.3) for the 36 listeners who judged the speech of children with cleft palate and 92.1% (11.1/12, *SD* = 0.3) for the 36 listeners who judged the speech of children without cleft palate. Mean intrajudge agreement ranged from 58.3% to 94.4% for the listeners for children with cleft palate and from 88.9% to 97.3% for the listeners for the children without cleft palate.

For the 42 listeners whose responses were collected using SIP-CCLP Ver. 3, it was possible to compare listeners who differed in the amount of experience in judging disordered speech. There were three levels of experience: second year speech-pathology graduate students, first year speech pathology graduate students and other students in rehabilitation medicine, with 14 listeners in each level. Intrajudge agreement was not significantly different among the three groups of listeners ($F_{(2,26)} = .44$, $p = .65$). When examined by group, mean intrajudge agreement was 87.5% (10.5/12, *SD* = 1.5) for the second year speech-pathology graduate students, 83.9% (10.1/12, *SD* = 1.6) for the first year speech-pathology graduate students and 87.5% (10.5/12, *SD* = 1.8) for the other students in rehabilitation medicine.

Interjudge reliability for the three listeners' intelligibility scores for each of the 24 children was examined by calculating an intraclass correlation coefficient (ICC). The ICC was 0.97 for the 24 groups of three listeners. By group, the ICC was 0.95 for the 12 groups of three listeners for the children with cleft palate and 0.96 for the 12 groups of three listeners for the children without cleft palate.

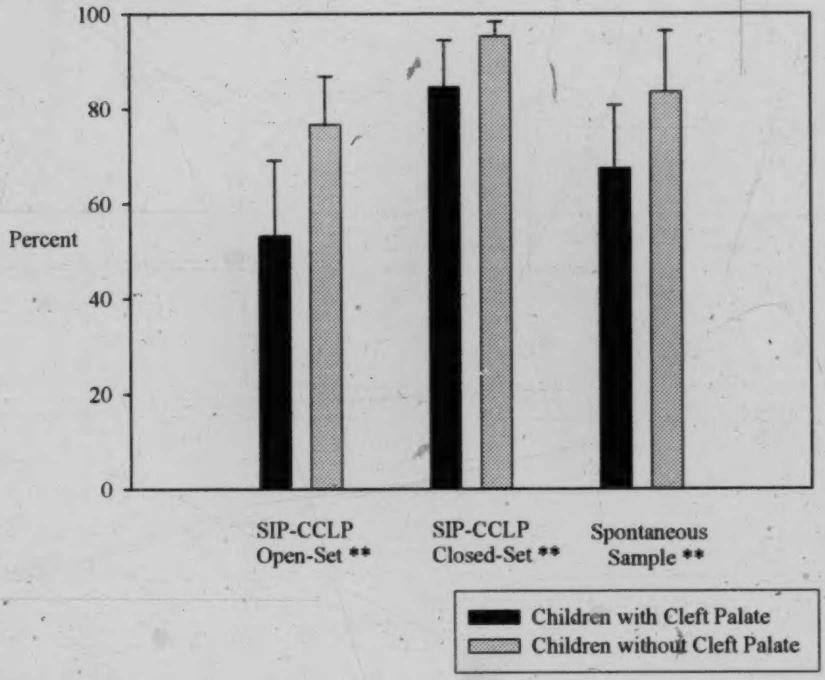
SIP-CCLP closed-set response task.

Intelligibility scores on the closed-set word response task ranged from 65.5% to 98.7% for the children with cleft palate (*M* = 84.5%, *SD* = 9.9) and from 90.3% to 99.0% for the children without cleft palate (*M* = 95.3%, *SD* = 3.0). Intelligibility scores were significantly lower for the children with cleft palate ($t = -3.6$, $p = 0.0015$). The effect size was 1.04. These scores are also shown in Figure 5.

Calculation of intrajudge agreement for the closed-set response task was possible for the listener judges for seven children with cleft palate (21 listeners) and seven children without cleft palate (21 listeners) because a software update was available to collect this information. Mean intrajudge agreement, based on the 20 repeated items, was 93.1% (18.6/20; *SD* = 6.3) for the 42 listeners. When examined by group, mean intrajudge agreement was 90.0% (18/20; *SD* = 7.8) for the listeners for the seven children with cleft palate and 96.2% (19.2/20; *SD* = 1.6) for the listeners for the seven children without cleft palate. Mean intrajudge agreement ranged from 78.3% to 96.7% for the listeners for the children with cleft palate and from 95% to 98.3% for the listeners for the children without cleft palate. Intrajudge agreement was not significantly different among the three groups of listeners (second year speech-pathology graduate students, first year speech pathology graduate students, other students in rehabilitation medicine) (F(2,26) = .042, p = .96). When examined by group, mean intrajudge agreement was 92.9% (18.6/20, *SD* = 9.8) for the second year speech-pathology graduate students, 93.6% (18.7/20, *SD* = 7.2) for the first year speech-pathology graduate students and 92.9% (18.6/20, *SD* = 9.1) for the other students in rehabilitation medicine.

Interjudge reliability for the three listeners' intelligibility scores for each of the 24 children was examined by calculating an intraclass correlation coefficient (ICC). The ICC was 0.92 for the 24 groups of three listeners. By group, the ICC was 0.88 for the 12 groups of three listeners for the children with cleft palate and 0.82 for the 12 groups of three listeners for the children without cleft palate.

Figure 5. Intelligibility scores for children with and without cleft palate for the SIP-CCLP open-set, SIP-CCLP closed-set and spontaneous sample open-set response tasks.

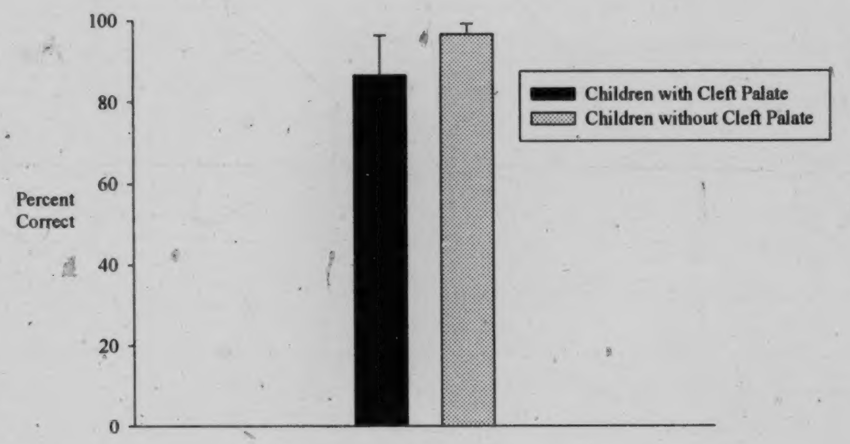


Note. **p<.01, one-tailed test.

Error Analysis: Closed-set Response Task

A total of 312 errors on phonetic contrast items were identified for the 12 children with cleft palate, compared to 79 for the children without cleft palate. Number of errors ranged from 0 to 65 for the 12 children with cleft palate (*M* = 25.8, *SD* = 18.9) and from 1 to 15 for the 12 children without cleft palate (*M* = 6.6, *SD* = 4.8). Mean percent phonetic contrast items correct was 86.5% (*SD* = 9.8) for the children with cleft palate and 96.7% (*SD* = 2.6) for the children without cleft palate. Mean percent phonetic contrast items correct was significantly lower for the children with cleft palate (t = -3.48, p = .002) with an effect size of 1.0. As expected, intelligibility scores on the closed-set response task (mean of three listeners) and percent phonetic contrast items correct (agreement of 2/3 listeners) were significantly correlated (r = .99, p < 0). For the correct/incorrect judgments, mean interjudge agreement was 86.3% (*SD* = 10.4). Mean interjudge agreement for the children with cleft palate was 80.3% (*SD* = 10.8%) and ranged from 64.3% and 95.6%. Mean interjudge agreement for the children without cleft palate was 92.3% (*SD* = 5.6%) and ranged from 83.0% to 98.5%. Figure 6 displays mean percent phonetic contrast items correct for the two groups of children.

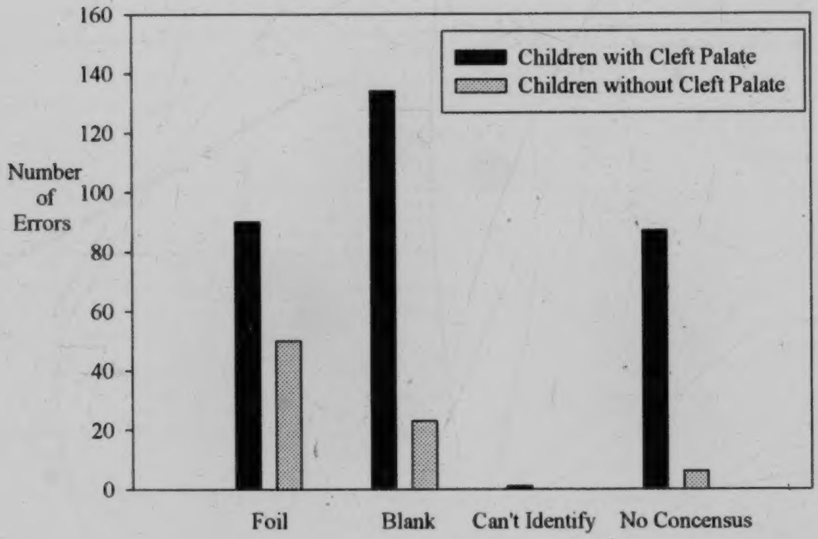
Figure 6. Percent phonetic contrast items correct by group.



Note. **p<.01, one-tailed test.

A minimum of 2 of 3 listeners chose the foil for 90 phonetic contrast items identified as errors for the children with cleft palate, and for 50 items for the children without cleft palate. Two of three listeners typed the same response in the "blank" for 134 errors for the children with cleft palate and for 23 errors for the children without cleft palate. Two of three listeners chose "can't identify" for 1 error for the children with cleft palate. For 87 errors for the children with cleft palate and for 6 errors for the children with typical speech, no consensus was reached among listeners on the error response. Breakdown of errors into these four categories is shown in Figure 7.

Figure 7. Frequency of errors by listener response for each group of children.



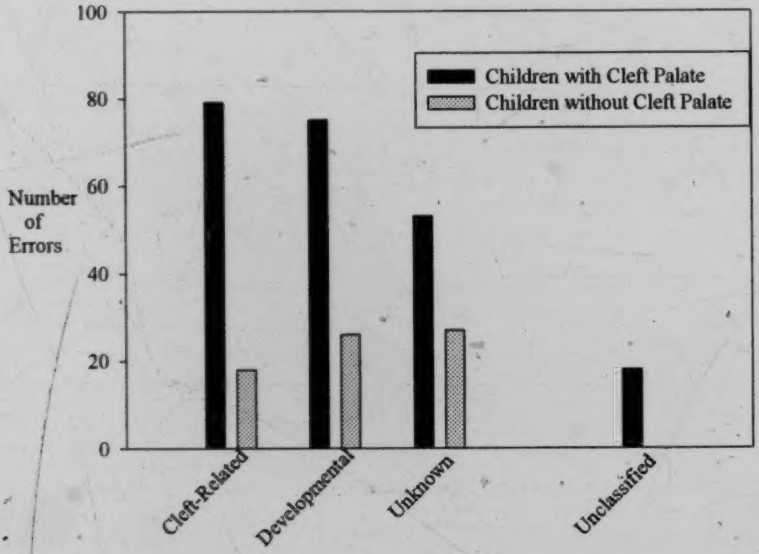
When listeners used the "blank" button to type in a sound in the target position that was not presented in either of the first two buttons, listeners identified additional errors that were not specifically tested in SIP-CCLP. For example, listeners typed in "geep" when presented with the contrast pair "cheep-jeep" where "jeep" was the target, indicating a cleft-related place preference error. These listener-generated errors were included in the calculations of the number of errors in each category.

A total of 79 errors classified as "cleft-related" were identified for the 12 children with cleft palate, compared to 18 for the 12 children without cleft palate. Number of cleft-related errors ranged from 0 to 20 for the children with cleft palate (*M* = 6.7, *SD* = 7.6) and from 0 to 4 for the children without cleft palate (*M* = 1.6, *SD* = 1.6). Contrary to expectations, number of cleft-related errors was not found to be significantly higher for the group of children with cleft palate (U = 43.0, p = .101). A total of 75 errors were classified as "developmental" for the children with cleft palate, compared to 26 for the children without cleft palate. Number of developmental errors ranged from 0 to 35 for the children with cleft palate (*M* = 5.83, *SD* = 9.6) and from 0 to 5 for the children without cleft palate (*M* = 2.2, *SD* = 1.7). Number of developmental errors was not significantly different between the two groups of children (U = 52.5, p = 0.266). A total of 53 errors were classified as "unknown" for the children with cleft palate, compared to 27 for the children without cleft palate. Number of "unknown" errors ranged from 0 to 10 for the children with cleft palate and from 0 to 8 for the children without cleft palate. Frequency of errors by category is shown in Figure 8.

For the children with cleft palate, a total of 18 errors were not able to be classified using the initial categories proposed for the study (i.e., cleft-related, developmental, unknown). These errors were considered as "unclassified". All errors identified for the children without cleft palate were able to be classified. Number of "unclassified" errors ranged from 0 to 8 for the children with cleft palate. These errors included three instances of a fricative target identified as a stop, two instances where a fricative target was identified as an affricate, one instance where a stop was identified as a fricative, five instances where a affricate was identified as an fricative, two instances where a stop was identified as a consonant cluster, two instances where a consonant cluster was identified as a fricative and three sonorant errors. The unclassified error patterns are listed in

Appendix M. Although some of the errors fit developmental patterns (i.e., "veil" to "dale" [stopping], "strip" to "frip" [cluster reduction]), these errors involve at least 2 features (e.g., voice, manner, place) and so were assigned to the "unclassified" category.

Figure 8. Frequency of errors by category and group.



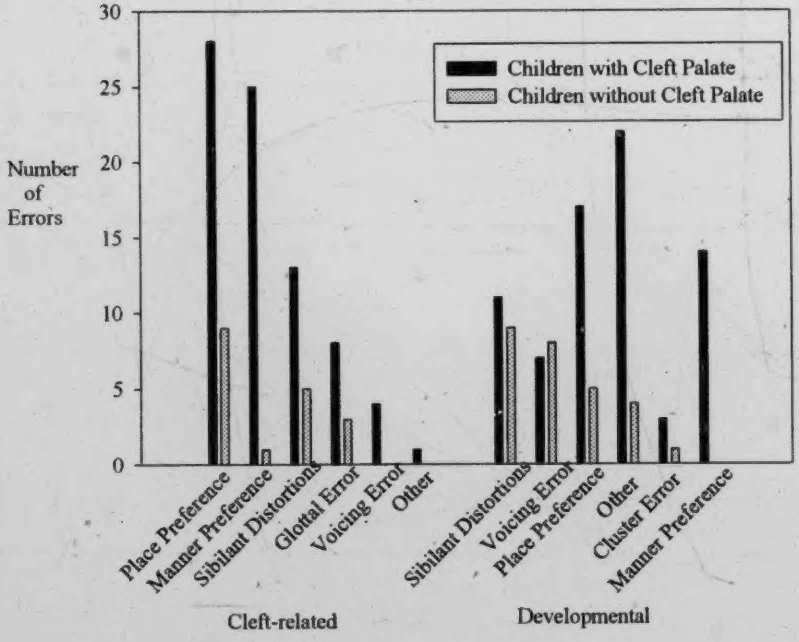
Frequency of the cleft-related, developmental and unknown errors by the six subcategories tested by SIP-CCLP is reported in Table 8 for the two groups of children. Figure 9 shows the frequency of cleft-related and developmental errors by subcategory. For the children with cleft palate, four cleft-related errors identified by listeners using the "blank" could not be classified as one of the six apriori subcategories. These errors were placed in the category "other" and include four instances where a fricative was identified as /sn/. Similarly, twenty-two developmental errors for the children with cleft palate and four developmental errors for the children without cleft palate were placed in the category "other". These errors include 18 instances where a liquid (i.e., /r/) was identified as a glide (i.e., /w/), four instances of an assimilation error where a glide (i.e., /j/) was identified as a liquid (i.e., /l/ in "yell"), three instances of deaffrication (i.e., /tʃ/ identified as /ʃ/) and one instance of a consonant cluster (i.e., /tr/) identified as an affricate (i.e., /tʃ/). A complete list of the cleft-related, developmental and unexpected errors identified for the children with cleft palate and their age-similar peers is included in Appendix N.

Table 8.

Frequency of Cleft-related, Developmental and Unknown Errors by SIP-CCLP Error Subcategories

Error Category	Error Subcategory	Children with Cleft Palate	Children without Cleft Palate
Cleft-related	Place Preference Error	28	9
	Manner Preference Error	25	1
	Sibilant Error	13	5
	Glottal Error	8	3
	Voicing Error	1	0
	Other	4	0
Developmental	Place Preference Error	17	5
	Manner Preference Error	14	0
	Sibilant Error	11	9
	Voicing Error	7	8
	Cluster Error	3	1
	Other	22	4
Unknown	Place Preference Error	3	1
	Manner Preference Error	2	3
	Sibilant Error	7	2
	Glottal Error	0	2
	Voicing Error	36	15
	Cluster Error	5	4
Unclassified		18	0

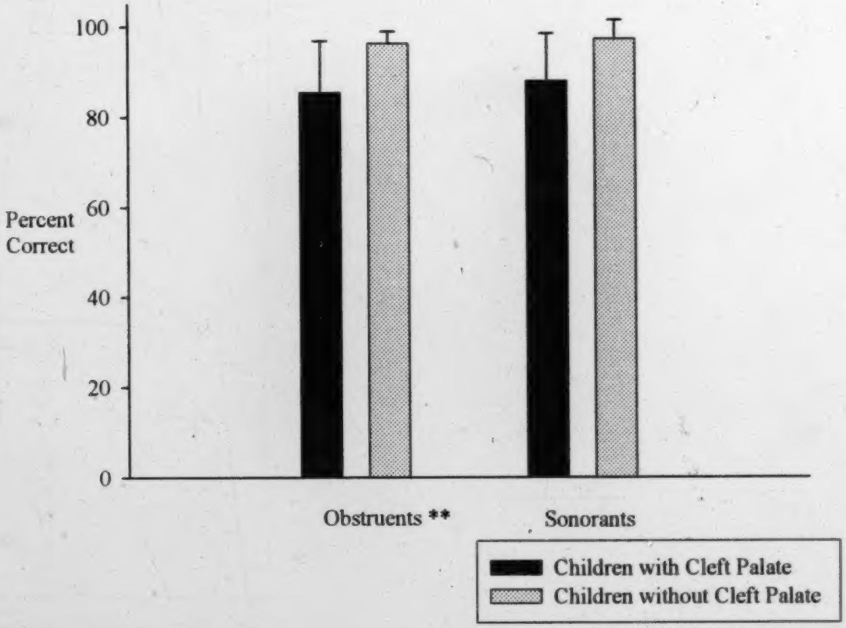
Figure 9. Contrast target error profile based on results from SIP-CCLP for the children with and without cleft palate.



Sound Analysis of Errors: Closed-set Response Task

Percent obstruents correct ranged from 63.1% to 100% for the children with cleft palate (*M* = 85.4, *SD* = 11.6) and from 90.4% to 99.3% for the children without cleft palate (*M* = 96.3, *SD* = 2.7). As hypothesized, percent obstruents correct was significantly lower for the children with cleft palate (t = -3.2, p = .004). The effect size was 0.92. Percent sonorants correct ranged from 70.4% to 100% for the children with cleft palate (*M* = 88.0, *SD* = 10.4) and from 88.6% to 100% for the children without cleft palate (*M* = 97.3, *SD* = 4.2). Information on sonorants is provided for interest only. Percent obstruents and sonorants correct for the two groups of children are shown in Figure 10.

Figure 10. Percent target obstruents and sonorants correct in the SIP-CCLP closed-set response task for children with and without cleft palate.

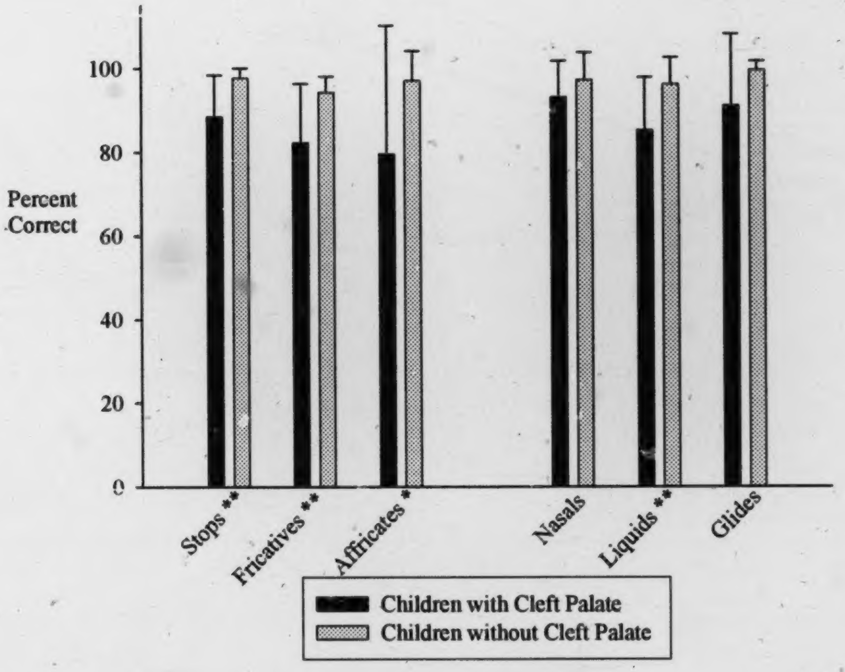


Note. **p<.01, one-tailed test.

Sound analysis of errors organized by manner.

For the children with cleft palate, percent stops, fricatives and affricates correct were 88.4% (*SD* = 10.0), 82.3% (*SD* = 14.1) and 79.5% (*SD* = 30.6), respectively. For the children without cleft palate, percent stops, fricatives and affricates correct were 97.6% (*SD* = 2.4), 94.3% (*SD* = 3.8) and 97.0% (*SD* = 7.1), respectively. Percent stops correct, percent fricatives correct and percent affricates correct were found to be significantly lower for the children with cleft palate (t = - 3.1, p = .0045; t = -2.9, p = .007; t = 1.9, p = .039 respectively). Percent liquids correct was 85.2% (*SD* = 12.6) for the children with cleft palate and 96.1% (*SD* = 6.4) for the children without cleft palate and was significantly lower for the children with cleft palate (t = -2.7, p = .008). Significant group differences were not found for percent nasals and glides correct for the two groups of children. Percent target sounds correct organized by manner are shown in Figure 11 for the two groups of children.

Figure 11. Percent target sounds correct organized by manner in SIP-CCLP closed-set response task for children with and without cleft palate.

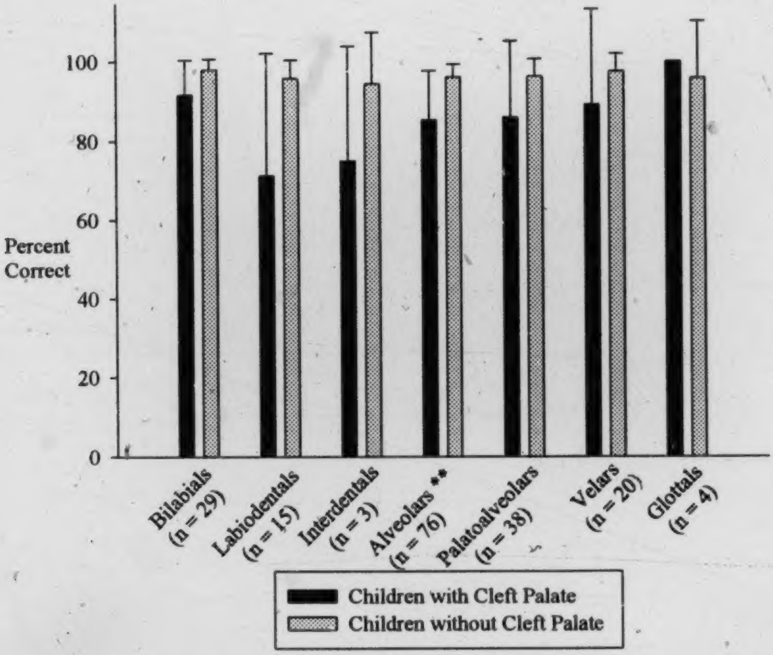


Note. * p<.05, one-tailed test. **p<.01, one-tailed test.

Sound analysis of errors organized by place of articulation.

Percent alveolars correct was 85.3% (*SD* = 12.4) for the children with cleft palate and 96.1% (*SD* = 3.2) for the children without cleft palate and was significantly different between the two groups (t = -2.9, p =.006). No significant differences were found between the two groups for percent bilabials, labiodentals, interdentals, palatoalveolars, velars or glottals correct. Percent target sounds correct organized by place of articulation is shown in Figure 12.

Figure 12. Percent target sounds correct organized by place of articulation in SIP-CCLP closed-set response task for children with and without cleft palate.

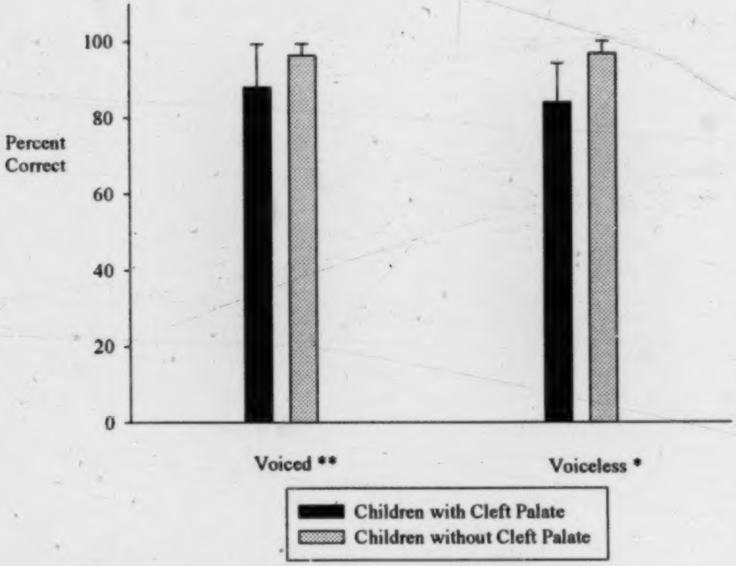


Note. No standard deviation is shown for glottal place of articulation for the children with cleft palate as all children achieved 100% correct for this place of articulation. **p<.01, one-tailed test.

Sound analysis of errors organized by voicing.

Percent voiced and voiceless consonants correct were 84.0% (*SD* = 10.2) and 87.9% (*SD* = 11.3) for the children with cleft palate and 96.7% (*SD* = 3.2) and 96.3% (*SD* = 3.0) for the children without cleft palate. Percent voiced and voiceless consonants correct were significantly lower for the children with cleft palate (voiced: t = -4.1, p<0.00; voiceless: t = -2.49, p = .014). Percent voiced and voiceless consonants correct were not significantly different for the children with cleft palate. Percent target sounds correct organized by voicing is shown in Figure 13 for the two groups of children.

Figure 13. Percent target sounds correct organized by voicing in SIP-CCLP closed-set response task for children with and without cleft palate.



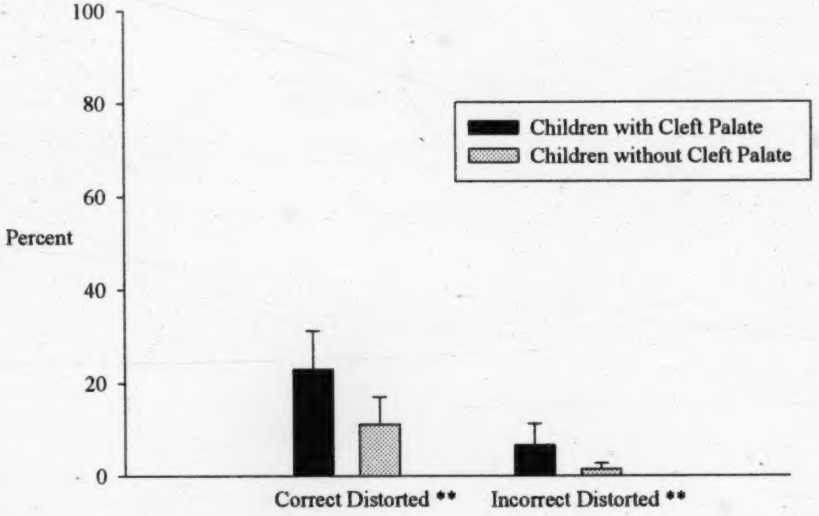
Note. **p<.01, one-tailed test. * p<.05, one-tailed test.

Distortion Analysis: Closed-set Response Task

Percent correct distorted items ranged from 11.7% to 34.8% (*M* = 23.0%, *SD* = 8.2) for the children with cleft palate and from 3.1% to 22.5% (*M* = 11.2%, *SD* = 5.8) for the children without cleft palate and were significantly higher for the children with cleft palate (t = 4.05, p = 0.001). The effect size was 1.17. Percent incorrect distorted items ranged from 0.9% to 16.7% for the children with cleft palate (*M* = 6.6%, *SD* = 4.7) and from 0% to 3.6% for the children without cleft palate (*M* = 1.5%, *SD* = 1.2) and were significantly higher for the children with cleft palate (t = 3.64, p = 0.0015). Effect size was 1.05. Percent correct and incorrect distorted scores for the two groups of children are shown in Figure 14.

Calculation of intrajudge agreement for the closed-set response task was possible for the 42 listener judges for seven children with cleft palate and seven children without cleft palate. Mean intrajudge agreement for rating each item as clear or distorted, based on the 20 repeated items, was 87.9% (17.6/20, *SD* = 7.6) for the 42 listeners. When examined by group, mean intrajudge agreement was 82.6% (16.6/20; *SD* = 7.4) for the 21 listeners for the seven children with cleft palate and 92.9% (18.6/20; *SD* = 3.4) for the 21 listeners for the seven children without cleft palate. Mean intrajudge agreement ranged from 71.7% to 86.7% for the listeners for the children with cleft palate and from 88.3% to 98.3% for the listeners for the children without cleft palate. Intrajudge agreement was not significantly different among the three groups of listeners (14 second year speech-pathology graduate students, 14 first year speech pathology graduate students, 14 other students in rehabilitation medicine) (F(2,26) = .49, p = .62). When examined by group, mean intrajudge agreement was 88.6% (17.7/20, *SD* = 12.9) for the second year speech-pathology graduate students, 82.9% (16.6/20, *SD* = 21.1) for the first year speech-pathology graduate students and 82.4% (16.5/20, *SD* = 19.8) for the other students in rehabilitation medicine. Interjudge agreement for the three sets of listeners for the children with cleft palate was 56.9% (*SD* = 10.7%) and for the children without cleft palate was 77.5% (*SD* = 7.2%).

Figure 14. Percent correct distorted and incorrect distorted items on the SIP-CCLP closed-set response task for children with and without cleft palate.



Note. **p<.01, one-tailed test.

Comparison of Children without Cleft Palate by Age Group

Descriptive information for the 24 children without cleft palate used in the age groups analysis is included in Table 7. Of these 24 children, six were in each of the following four age groups: 36 – 47 months, 48 – 59 months, 60 – 71 months and 72 – 84 months. All children had typical speech and language development and no history of craniofacial abnormalities.

Intelligibility Analysis

SIP-CCLP open-set response task.

On the open-set response task, as expected, intelligibility scores among the four groups of children were significantly different ($F_{(3,20)} = 3.6$, $p = .032$). Post-hoc analysis (Tukey) revealed that intelligibility scores for the youngest children (36 – 47 months) (*M* = 72.9%, *SD* = 12.1) were significantly lower than the intelligibility scores for the oldest children (72 – 84 months) (*M* = 87.0, *SD* = 3.5) with an effect size of 1.12. No other significant group differences were found. Intelligibility scores ranged from 55.9% to 88.9% for the children 36 – 47 months, from 73.7% to 86.6% for the children 48 – 59 months, 71.6% to 88.1% for the children 60 – 71 months and 84.0% to 93.0% for the children 72 – 84 months. Intelligibility scores for the SIP-CCLP open-set response task for the four groups of children are shown in Table 9.

Table 9.

SIP-CCLP and Spontaneous Sample Scores for the Four Age Groups of Children without Cleft Palate

	Dependent Variable	36 – 47 months	48 – 59 months	60 – 71 months	72 – 84 months
SIP-CCLP open-set	Intelligibility Score (%)	72.9 (*SD* = 12.1)	81.0 (*SD* = 5.9)	81.6 (*SD* = 5.9)	87.0 (*SD* = 3.5)
	Intrajudge Reliability	11.1/12 (*SD* = 0.13)	11.3/12 (*SD* = 0.6)	10.8/12 (*SD* = 0.3)	11.2/12 (*SD* = 0.7)
SIP-CCLP closed-set	Intelligibility Score (%)	94.8 (*SD* = 3.3)	93.5 (*SD* = 4.4)	95.4 (*SD* = 2.2)	97.3 (*SD* = 1.5)
	Percent Phonetic Contrast Items Correct	96.0 (*SD* = 2.8)	94.4 (*SD* = 5.1)	96.5 (*SD* = 2.6)	98.3 (*SD* = 1.0)
Spontaneous sample	Intelligibility Score (%)	81.7 (*SD* = 13.7)	87.9 (*SD* = 6.4)	87.6 (*SD* = 6.8)	90.7 (*SD* = 2.9)

Mean intrajudge agreement, based on the 12 repeated items in the open-set response task, was 92.5% (11.1/12, *SD* = 0.5) for the 72 listeners. When examined by group, mean intrajudge agreement for the four age groups of children, in order of increasing age, was 92.5% (11.1/12, *SD* = 0.1), 94.2% (11.3/12, *SD* = 0.6), 90.0% (10.8/12, *SD* = 0.3) and 93.3% (11.2/12, *SD* = 0.7). Mean intrajudge agreement, in order of ascending age, ranged from 91.7% to 94.2% (36 – 47 months), 88.9% to 100% (48 – 59 months), 88.9% to 100% (60 – 71 months) and 86.1% to 100% (72 – 84 months) for the four groups of children.

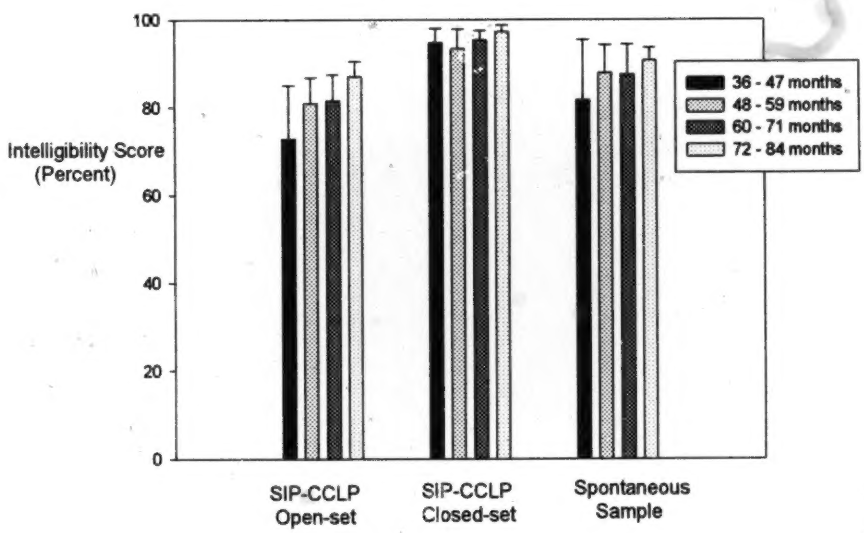
Interjudge reliability for the three listeners' intelligibility scores for each child was examined by calculating an intraclass correlation coefficient (ICC). The ICC was 0.95 for the 24 groups of three listeners. For the four groups of children, in order of increasing age, the ICC was 0.99 (6 groups of three listeners), 0.92 (6 groups of three listeners), 0.81 (6 groups of three listeners) and 0.66 (6 groups of three listeners).

SIP-CCLP closed-set response task.

Intelligibility scores on the closed-set response task were similar among the four groups of children ($F_{3,20} = 1.65$, $p = 0.209$). Intelligibility scores on the closed-set response task for the four age groups of children were as follows: 36 – 47 months 94.8% (*SD* = 3.3), 48 – 59 months 93.5% (*SD* = 4.4), 60 – 71 months 95.4% (*SD* = 2.2) and 72 – 84 months 97.3% (*SD* = 1.5). For the four groups of children, intelligibility scores ranged from 90.3% to 95.9% (36 – 47 months), 86.2% to 99.0% (48 – 59 months), 92.5% to 98.1% (60 – 71 months) and 94.5% to 99.1% (72 – 84 months). Intelligibility scores for the SIP-CCLP closed-set response task for the four age groups of children are provided in Table 9.

Calculation of intrajudge agreement for the closed-set response task was possible for 45 listener judges for four children 36 – 47 months, four children 48 – 59 months, three children 60 – 71 months and four children 72 – 81 months. Mean intrajudge agreement, based on the 20 repeated items, was 97.3% (19.5/20, *SD* = 0.4) for the 45 listeners. When examined by group, mean intrajudge agreement was 97.9% (19.6/20, *SD* = 0.5) for the 12 listeners for the children 36 – 47 months, 96.3% (19.3/20, *SD* = 0.5) for the 12 listeners for children 48 – 59 months, 97.8% (19.6/20, *SD* = 0.2) for the 12 listeners for children 60 - 71 months and 97.5% (19.5/20, *SD* = 0.3) for the 12 listeners for children 72 - 84 months. Figure 15 displays the intelligibility scores for the four age groups of children with typical speech on the SIP-CCLP open-set and SIP-CCLP closed-set response tasks.

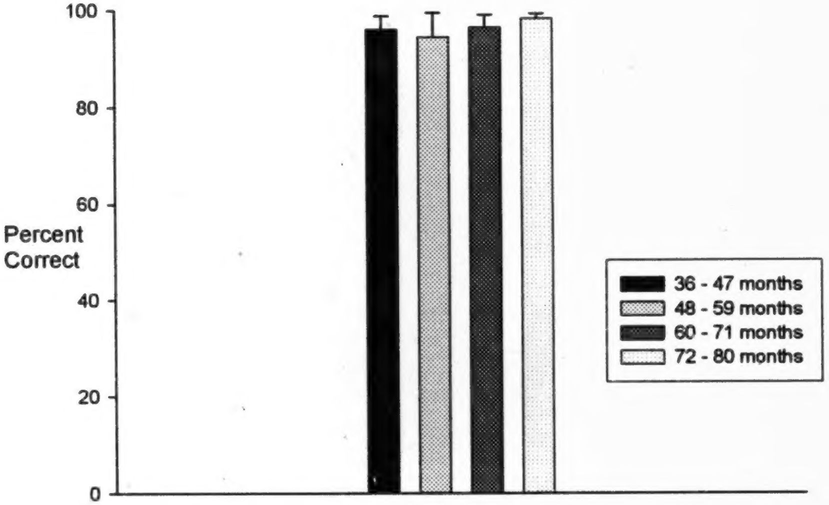
Figure 15. Intelligibility scores on the SIP-CCLP open-set and closed-set response tasks, compared with the spontaneous sample intelligibility scores for the four age groups of children without cleft palate.



Error Analysis: Closed-set Response Task

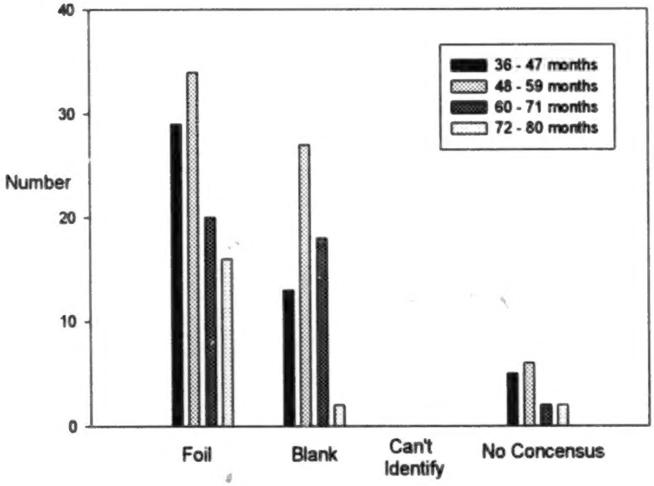
A total of 47 errors were identified for the children 36 – 47 months, compared to 66 errors for the children 48 – 59 months, 40 errors for the children 60 – 71 months and 20 errors for the children 72 – 84 months. Number of errors ranged from 1 to 15 for the youngest children (*M* = 7.8, *SD* = 5.0), from 1 to 29 for the children 48 – 59 months (*M* = 11.0, *SD* = 9.6), from 2 to 13 for the children 60 – 71 months (*M* = 6.7, *SD* = 4.8) and from 2 to 7 for the children 72 to 84 months (*M* = 3.3, *SD* = 2.0). Mean percent phonetic contrast items correct for the four age groups of children without cleft palate were as follows: 36 – 47 months 96.0% (*SD* = 2.8), 48 – 59 months 94.4% (*SD* = 5.1), 60 – 71 months 96.5% (*SD* = 2.6) and 72 – 84 months 98.3% (*SD* = 1.0). Contrary to expectations, mean percent phonetic contrast items correct was similar among the four groups of children ($F_{(3,20)} = 1.49$, $p = .248$). Intelligibility scores on the closed set response task (mean of three judges) and percent phonetic contrast items correct (based on 2/3 agreement) were significantly correlated ($r = .94$, $p < 0$). For the correct/incorrect judgments, interjudge agreement for four age groups of children without cleft palate were as follows: 36 – 47 months 90.6% (*SD* = 6.4%), 48 – 59 months 91.3% (*SD* = 5.1%), 60 – 71 months 93.1% (*SD* = 3.5%) and 72 – 84 months 95.1% (*SD* = 2.9%). Figure 16 displays mean percent phonetic contrast items correct for children by age group.

Figure 16. Percent phonetic contrast items correct for the four age groups of children without cleft palate.



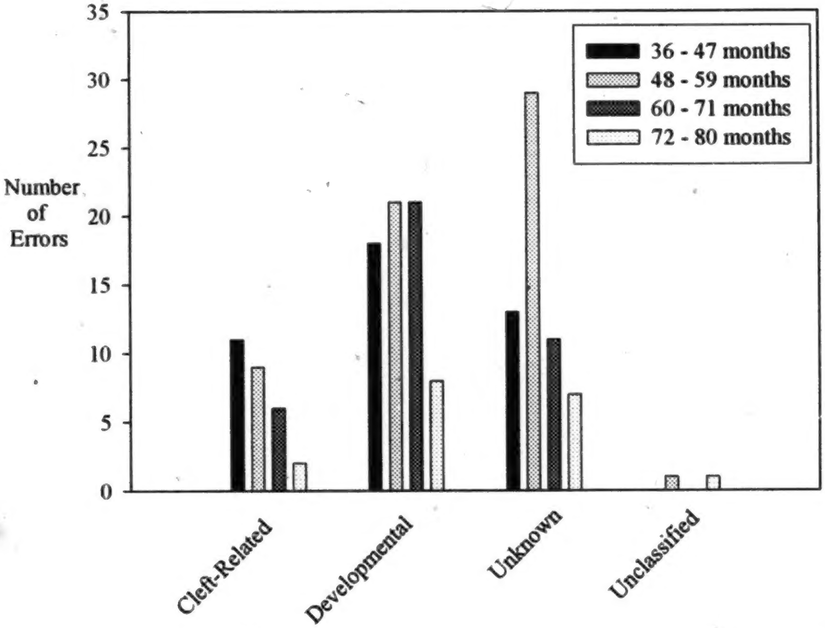
Of the 47 errors identified by listeners for the children 36 – 47 months, a minimum of 2 of 3 listeners either chose or typed in the same response for 42 errors. For the children 48 – 59 months, listener agreement was obtained on 60 of the 66 errors. For the children 60 - 71 months, listener agreement was obtained on 38 of the 40 errors. For the children 72 – 84 months, listener agreement was obtained on 18 of the 20 errors. A minimum of 2 of 3 listeners chose the foil for 29 contrast items for the children 36 – 47 months, for 33 contrast items for the children 48 – 59 months, for 20 contrast items for the children 60 – 71 months and for 16 contrast items for the children 72 – 84 months. The majority of listeners typed the same response in the "blank" for 13 errors for the children 36 – 47 months, for 27 errors for the children 48 – 59 months, for 18 errors for the children 60 – 71 months and for 2 errors for the children 72 – 84 months. No consensus was reached for 5 errors for the children 36 – 47 months, for 6 errors for the children 48 – 59 months, for 2 errors for the children 60 – 71 months and for 2 errors for the children 72 – 84 months. Frequency of errors in these four categories is shown in Figure 17.

Figure 17. Frequency of errors by listener response for the four age groups of children without cleft palate.



A total of 12 errors classified as "cleft-related" were identified for the children 36 – 47 months, compared to 8 for the children 48 – 59 months, 6 for the children 60 – 71 months and 2 for the children 72 – 84 months. Number of cleft-related errors ranged from 0 to 4 for the children 36 – 47 months (*M* = 1.8, *SD* = 1.8), from 0 to 5 for the children 48 – 59 months (*M* = 1.5, *SD* = 1.9), from 0 to 2 for the children 60 – 71 months (*M* = 1.0, *SD* = 0.9) and from 0 to 2 for the children 72 – 84 months (*M* = 0.3, *SD* = 0.8). A total of 18 errors were classified as "developmental" for the children 36 – 47 months, compared to 21 for the children 48 – 59 months, 21 for the children 60 – 71 months and 9 for the children 72 – 84 months. Number of developmental errors ranged from 1 to 5 for the children 36 – 47 months, from 0 to 8 for the children 48 – 59 months, from 1 to 11 for the children 60 – 71 months and from 1 to 2 for the children 72 – 84 months. Frequency of errors by category is shown in Figure 18. A total of 14 errors were classified as "unknown" for the children 36 – 47 months, compared to 28 for the children 48 – 59 months, 11 for the children 60 – 71 months and 7 for the children 72 – 84 months. Number of "unknown" errors ranged from 0 to 8 for the children 36 – 47 months, from 0 to 14 for the children 48 – 59 months, from 0 to 5 for the children 60 – 71 months and from 0 to 2 for the children 72 – 84 months. Table 10 contains the breakdown by subcategory for cleft-related, developmental and unknown categories for the four age groups of children. A complete list of all the errors identified for the children, organized by age group, is included in Appendix O.

Figure 18. Frequency of errors by category (cleft-related, developmental, unknown and unclassified) and age group for the children without cleft palate.



Two errors were not able to be classified as cleft-related, developmental or unknown. These errors were considered as "unclassified". Errors that occurred that were not included in the initial categories analyzed in the SIP-CCLP were one instance where a stop was identified as a nasal/stop cluster (i.e., "lap" was heard as "lamp") and one instance where a glide was identified as a liquid (i.e., "walk" was heard as "rock").

Table 10.

Frequency of Cleft-related, Developmental and Unknown Errors by SIP-CCLP Error Subcategories and by Group.

Error Category	Error Subcategory	36 – 47 months	48 – 59 months	60 – 71 months	72 – 84 months
Cleft-related	Place Preference Error	7	2	1	0
	Manner Preference Error	1	1	1	0
	Sibilant Error	1	3	2	2
	Glottal Error	2	3	1	0
	Voicing Error	0	0	1	0
Developmental	Place Preference Error	5	1	1	1
	Manner Preference Error	0	1	5	1
	Sibilant Error	8	7	0	1
	Voicing Error	2	8	4	5
	Cluster Error	2	0	2	0
	Other	1	4	9	0
Unknown	Place Preference Error	0	1	0	1
	Manner Preference Error	0	3	2	0
	Sibilant Error	2	0	0	1
	Glottal Error	2	0	0	0
	Voicing Error	6	16	5	3
	Cluster Error	3	9	3	3
Unclassified		0	1	0	1

Criterion Validity

Intelligibility

Comparison of Children with and without Cleft Palate

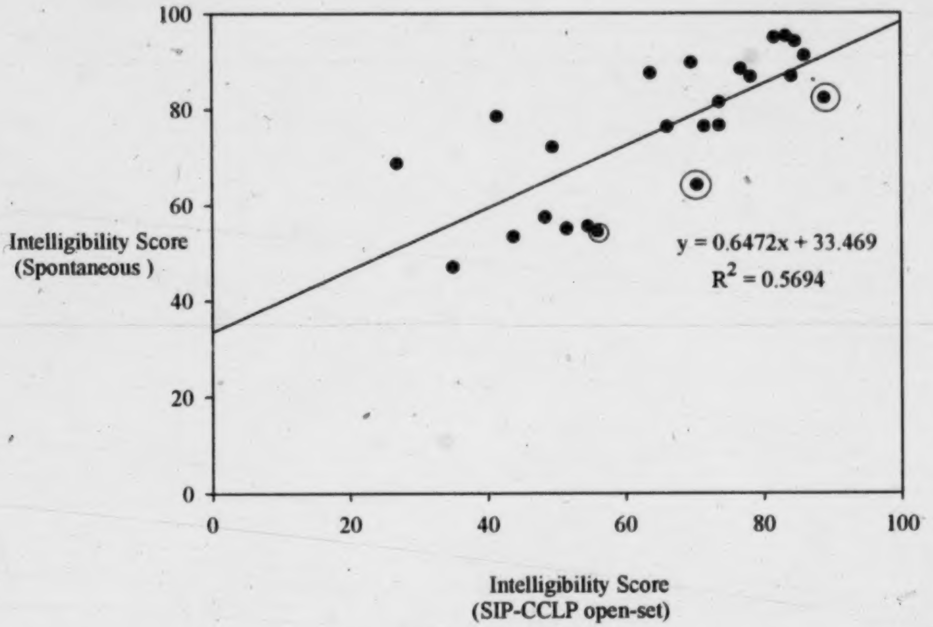
The group mean intelligibility score on the spontaneous sample was significantly higher for the 12 age-similar children without cleft palate than for the children with cleft palate (t = -3.03, p = 0.003). Spontaneous speech intelligibility scores ranged between

47.0% and 88.3% for the 12 children with cleft palate (*M* = 67.5%, *SD* = 13.3) and from 54.4% and 95.2% for the 12 children without cleft palate (*M* = 83.6%, *SD* = 12.7). As shown in Figure 19, all children, with the exception of three children without cleft palate (circled), had higher scores for the spontaneous speech sample than for the SIP-CCLP items. For the children with cleft palate, this difference ranged from 0.9% (CP12) to 41.8% (CP10) with a mean of 14.2% (*SD* = 13.0). For the children without cleft palate, this difference ranged from 1.6% (TS01) to 23.8% (TS03) with a mean of 9.3% (*SD* = 6.9).

Interjudge reliability for the three listeners' intelligibility scores for the spontaneous speech samples for each of the 24 children was examined by calculating an intraclass correlation coefficient (ICC). The ICC was 0.93 for the 24 groups of three listeners. By group, the ICC was 0.87 for the 12 groups of three listeners for the children with cleft palate and 0.93 for the 12 groups of three listeners for the children without cleft palate.

A moderately strong positive correlation was found between intelligibility scores on the SIP-CCLP and the spontaneous sample ($r = 0.76$, $p < 0.00$) for the 24 children which was lower than Gotzke (2003) but similar to Gordon-Brannan and Hodson (2000). When examined by group, a moderate positive correlation was found between intelligibility scores on the SIP-CCLP and the spontaneous sample for each of the two groups of children (children with cleft palate: $r = 0.62$, $p = .016$; children without cleft palate: $r = 0.67$, $p = .008$). Figures 20 and 21 show the relationship between SIP-CCLP open-set and spontaneous sample intelligibility scores for each group of children.

Figure 19. Relationship between SIP-CCLP open-set and spontaneous sample intelligibility scores for children with and without cleft palate.



Note. Circled points are those children whose SIP-CCLP intelligibility scores were higher than their spontaneous sample intelligibility scores.

Figure 20. Relationship between SIP-CCLP open-set and spontaneous sample intelligibility scores for children with cleft palate.

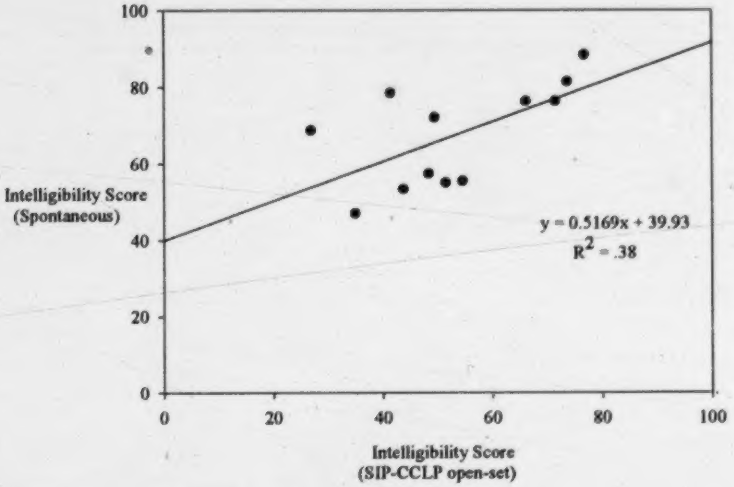
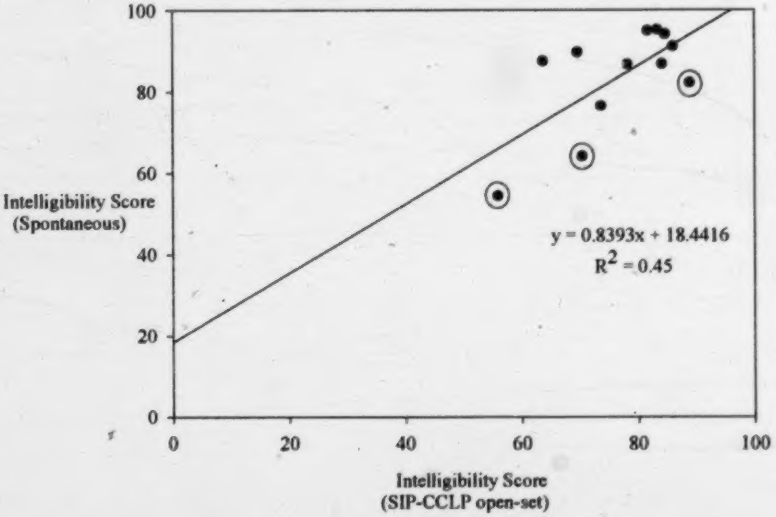


Figure 21. Relationship between SIP-CCLP open-set and spontaneous sample intelligibility scores for children without cleft palate.



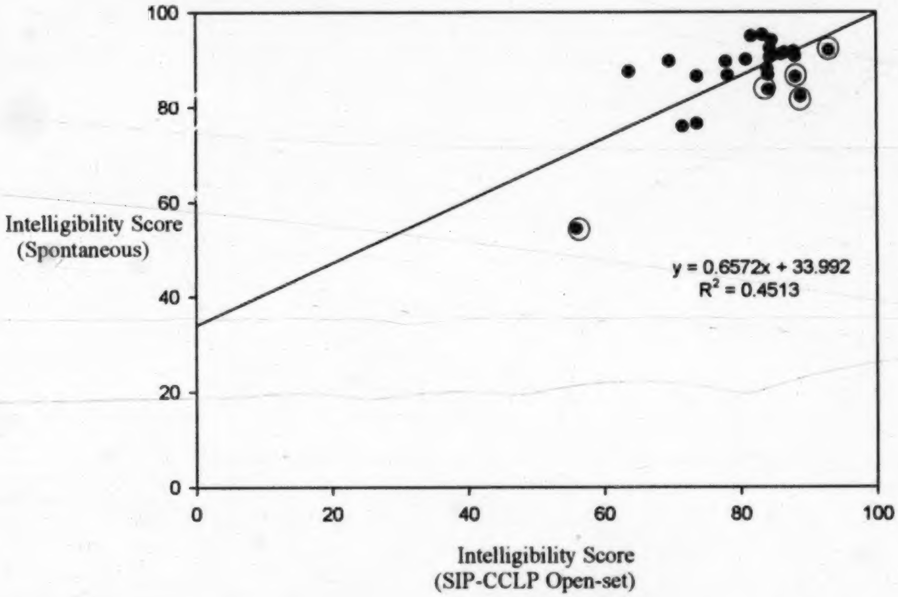
Note. Circled points are those children whose SIP-CCLP intelligibility scores were higher than their spontaneous sample intelligibility scores.

Comparison of Children without Cleft Palate by Age Group

The intelligibility scores on the spontaneous sample for the 4 age groups of children without cleft palate were as follows: 36 – 47 months 81.7% (*SD* = 13.7), 48 – 59 months 87.9% (*SD* = 6.4), 60 – 71 months 87.6% (*SD* = 6.8) and 72 – 84 months 90.7% (*SD* = 2.9). Figure 15 displays the spontaneous sample intelligibility scores for the four groups of children.

The mean intelligibility scores on the spontaneous sample were similar among the four age groups of children ($F_{(3,20)} = 1.21$, $p = 0.333$). As shown in Figure 22, all but five of the 24 children (circled) had higher scores for the spontaneous speech sample than for the SIP-CCLP open-set response task. A significant moderate positive correlation was found between intelligibility scores on the SIP-CCLP and the spontaneous sample ($r = 0.67$, $p < 0.00$).

Figure 22. Relationship between SIP-CCLP open-set and spontaneous sample open-set intelligibility scores for 24 children without cleft palate.



Note. Circled points are those children whose SIP-CCLP intelligibility scores were higher than their spontaneous sample intelligibility scores.

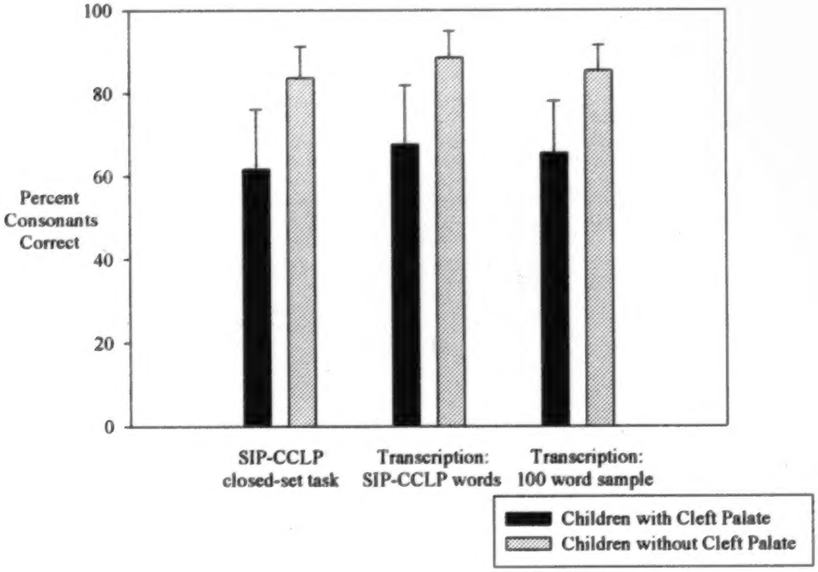
Phonetic Analysis

Children with and without Cleft Palate

For the children with cleft palate, the mean percent consonants correct (PCC) in the SIP-CCLP closed-set response task, the phonetic transcription of the SIP-CCLP words and phonetic transcription of the 100-word spontaneous sample were 61.7% (*SD* = 14.5), 67.6% (*SD* = 14.4) and 65.4% (*SD* = 12.7), respectively. No significant difference was found among the three sampling conditions for the children with cleft palate ($F_{(2, 22)}$ = 1.61, p = 0.225). Figure 23 displays the mean PCC for the three conditions for the children with cleft palate.

For the 12 age-similar children without cleft palate, the PCC in the SIP-CCLP closed-set response task, the phonetic transcription of the SIP-CCLP words and phonetic transcription of the 100-word spontaneous sample were 83.7% (*SD* = 7.6), 88.6% (*SD* = 6.3) and 85.4% (*SD* = 6.1), respectively. A significant difference was found among the three conditions ($F_{(2,22)} = 9.5$, $p = .001$). Post-hoc analyses revealed a significant difference between the closed-set response task and the phonetic transcription of the SIP-CCLP words ($F = 14.9$, $p = .003$) and between the phonetic transcription of the SIP-CCLP words and the spontaneous sample ($F = 17.4$, $p = .002$). Between the closed-set response task and the SIP-CCLP word transcription, this difference ranged from 0.7% to 14.6% with a mean of 5.2% (*SD* = 4.2). Between the phonetic transcription of the SIP-CCLP words and the spontaneous sample, this difference ranged from 0% to 7.3% with a mean of 3.7% (*SD* = 2.0). When analyzed by manner, percent stops correct was significantly higher in the phonetic transcription of the SIP-CCLP words than in the phonetic transcription of the spontaneous sample ($t = 4.66$, $p < .000$). Mean percent stops correct was 95.6% (*SD* = 4.6) for the SIP-CCLP words and 84.9% (*SD* = 8.7) for the spontaneous sample. Figure 23 displays the mean PCC for the three conditions for the children without cleft palate.

Figure 23. Percent consonants correct for children with and without cleft palate for the SIP-CCLP closed-set response task, the phonetic transcription of the SIP-CCLP words and the phonetic transcription of the 100-word spontaneous sample.



DISCUSSION

The purpose of this study was to conduct a further evaluation of the validity and reliability of the Speech Intelligibility Probe for Children with Cleft Palate Version 3 (SIP-CCLP Ver. 3) as a measure of intelligibility. Interjudge and intrajudge reliability of the SIP-CCLP was evaluated for listener judgments of correct/incorrect (open-set and closed-set response tasks) and ratings of correct/distorted (closed-set response task). Construct validity was evaluated by comparing scores obtained from the open-set and closed-set response tasks for children with and without cleft palate and for children without cleft palate of different ages. Criterion validity of the SIP-CCLP was evaluated by correlating intelligibility scores obtained from the SIP-CCLP and a spontaneous speech sample using open-set response tasks. Criterion validity was also evaluated by comparing speech sound error patterns on SIP-CCLP and those in the spontaneous speech sample.

Reliability

Identification Tasks

As predicted, intrajudge agreement on the repeated items in the open-set and closed-set response tasks was greater than 80%. However, intrajudge agreement was lower for the group of children with cleft palate compared to the group of children without cleft palate. The nonstandard substitutions and distortions present in the speech of children with cleft palate may make it more difficult for listeners to identify words and sounds. Intrajudge agreement was higher for the closed-set response task than the open-set response task. This result may be related to facilitating effect of context. In the closed-set response task, listeners were provided with two minimally contrastive words, which may have minimized the distracting effects of speech sound errors on sounds other than the target.

As predicted, a strong positive correlation was found for interjudge reliability on the SIP-CCLP open-set and closed-set response tasks. However, for the closed-set response task, interjudge reliability for the children without cleft palate was lower than for the children with cleft palate. When data for the children were examined, for one child without cleft palate, the difference between the highest and lowest score across the three judges was greater than 10%. When this child was eliminated from the ICC calculation, the ICC for the 11 groups of three listeners increased to 0.91, which is comparable to the ICC obtained for the open-set response task. Of note is that intrajudge agreement was not significantly different among the three groups of listeners (second-year speech pathology students, first-year speech pathology graduate students and other students in rehabilitation medicine) despite differing amounts of training and experience in judging disordered speech. This result suggests that training and experience of listeners on judging disordered speech does not affect their judgments when education level, language background, hearing status and motivation are similar between groups.

For the comparison of children without cleft palate of different ages, interjudge reliability decreased as the age of the groups increased for the SIP-CCLP open-set intelligibility scores. When data for the children were examined, for one child 60 - 71 months and one child 72 - 84 months, the difference between the lowest open-set intelligibility score and the highest score across the three judges was greater than 10%. When data for these two children are removed, the ICC increased to 0.91 for the children 60 - 71 months and to 0.81 for the children 72 – 84 months. Individual listener characteristics including attention, physical comfort, understanding and interest may have affected their scores. Furthermore, because of the large number of listening sessions that were completed, some listeners had more experience with the task. Yorkston and Beukelman (1980) found that listeners' open-set intelligibility scores were significantly higher after multiple exposures to the same word set. Further research is necessary to determine how increased number of exposures to the same task affected SIP-CCLP listener responses.

Ratings of Distorted/Clear

For the ratings of "distorted/clear" in the SIP-CCLP closed-set response task, intrajudge agreement was above 80% for both groups of children. However, interjudge agreement for the ratings was considerably below this value for the groups of listeners for the children with and without cleft palate (56.9% and 77.5%, respectively). This result suggests that different judges are less reliable in judging distortions than they are in identifying phonemes, and follows what has been found for transcription tasks. Shriberg and Lof (1991) reported higher interjudge agreement for broad transcription (e.g., "correct/incorrect" responses) than for narrow transcription (e.g., "distorted/clear" ratings).

Intrajudge agreement for "accuracy" scores, in which both "correctness" and "distorted/clear" ratings were considered, were lower than when these two variables were examined separately. The children with cleft palate had lower intrajudge agreement than the children without cleft palate, which is the same pattern seen for the judgments of "correct/incorrect" and for the ratings of "clear/distorted". As the speech of children with cleft palate may contain more sound errors and more distortions than children without cleft palate, lower reliability may be expected.

Conclusions

Overall, intra and inter-judge agreement results obtained for this study are interpreted as support for the reliability of the SIP-CCLP. Test-retest reliability was not conducted in this assessment of the SIP-CCLP; therefore, it is not possible to provide information about standard error of measurement and confidence intervals forchildren's "true scores". It is recommended that test-retest reliability be addressed in the next stage of the SIP-CCLP evaluation.

Construct Validity

Comparison of Children with and without Cleft Palate

SIP-CCLP Open-set Response Task

As predicted, intelligibility scores on the SIP-CCLP open-set response task for the children with cleft palate were significantly lower than intelligibility scores for the children without cleft palate. Intelligibility scores on this task for the children with cleft palate ranged from 26.9% to 76.8%, similar to the range of intelligibility scores described by Whitehill and Chau (2004) (i.e., 32.3% to 86.7%) in their study of 15 Cantonese speakers (aged 5 to 44 years) with cleft lip and palate (bilateral or unilateral). In the Whitehill and Chau study, intelligibility scores were obtained by a closed-set response task in which listeners were given a choice of one target and three foils. Mean intrajudge agreement for the open-set intelligibility scores was greater than 80%, similar to that reported by Hodge (1996) for open-set intelligibility scores for young children with and without dysarthria. Intrajudge agreement was lower for the children with cleft palate than for the children without cleft palate. It might be expected that listeners would be less reliable in identifying words in speech that contains more sound errors and distortions and is more difficult to understand.

SIP-CCLP Closed-set Response Task

Intelligibility score and percent phonetic contrast items correct

As expected, the children with cleft palate had a lower mean intelligibility score and fewer phonetic contrast items correct on the SIP-CCLP closed-set response task compared to the children without cleft palate. The intelligibility score on the closed-set task was based on the average number of contrast items that listeners identified correctly, whereas the phonetic contrast items correct was based on the percentage of contrast items that a minimum of 2 of 3 listeners identified correctly. Despite the differences in how these values were calculated, they were strongly positively correlated. The difference between the means for the two sets of scores was 2.0% for the children with cleft palate and 1.4% for the children without cleft palate. Intrajudge reliability was higher for the SIP-CCLP closed-set intelligibility scores than for Whitehill & Chau's (2004) closed-set task in which listeners were given choices of one target and three foils.

The mean intelligibility score for the SIP-CCLP closed-set response task was higher than the mean intelligibility score obtained from the SIP-CCLP open-set response task. This result is comparable to those obtained by Yorkston and Beukelman (1978) who also found that intelligibility scores for single words obtained using a closed-set response task (choice of 10 words) were higher than those obtained using an open-set response task for eight speakers with dysarthria.

For the group of children with cleft palate, the range of intelligibility scores obtained using the SIP-CCLP closed-set response task (i.e., 65.5% to 98.7%) was higher than the range of intelligibility scores obtained by Whitehill and Chau (2004) in their closed-set response task (i.e., 32.3% to 86.7%). This result may be related to differences in the type of response choices. Although listeners were given four response choices in the SIP-CCLP closed-set response task and in the Whitehill and Chau study, in the former, these choices were a target, a foil, a "blank" and can't identify; whereas in the latter, choices were one target and three foils. Yorkston and Beukelman (1980) found that as the number of foils in a closed-set response task increased, intelligibility scores decreased. Increased number of foils may increase the difficulty of the listening task and account for the difference in range of intelligibility scores obtained in the two studies. Further research is necessary to determine what effect changing the number of foils would have on intelligibility scores obtained using the SIP-CCLP.

Error pattern analysis.

A greater number of errors were identified for the children with cleft palate than the children without cleft palate. As well, there were a greater number of errors for the children with cleft palate (n = 87) than for the children without cleft palate (n = 6) in which no consensus was reached among the three listeners. For the children with cleft palate, the combination of nonstandard distortions and substitutions, hypernasality and errors on sounds other than the target may have made it more difficult for listeners to identify the sound produced, thereby resulting in an increased number of errors in which listeners did not agree on the perceived sound error.

Listeners used the "blank" response option to type in an alternative sound for the majority of errors identified for the children with cleft palate. From these responses, additional instances of some of the error patterns on the SIP-CCLP were identified. For example, for the children with cleft palate, listener-generated errors accounted for 67.9% and 63.6% of the errors identified as cleft-related place preference and cleft-related manner preference errors, respectively. Responses entered into the "blank" option also served to identify some sound substitution patterns that were not specifically tested in SIP-CCLP. For example, as a result of listeners typing a response in the "blank", for the children with cleft palate, 14 instances of gliding (i.e., /r/ is heard as /w/) were identified. However, some of the error patterns identified by listeners using the "blank" option were not classifiable.

Although listeners identified 79 cleft-related errors for the group of children with cleft palate and 18 cleft-related errors for the group of children without cleft palate, the number of cleft-related errors identified for the children with cleft palate was not significantly different than the number identified for the children without cleft palate. This result is contrary to expectations and may in part be due to the characteristics of the children recruited to participate and the variability in the severity of their speech disorder. Despite having inclusion criteria that allowed recruitment of children with cleft palate and expressive language delay, all of the children recruited for this study had both receptive and expressive language skills within normal limits. Morris and Ozanne (2003), through broad phonetic transcription and computerized phonological process analysis of spontaneous speech and articulation test responses, found that 3 year-old children with cleft palate and delayed expressive language used a smaller number of phonemes productively, used glottal stops as compensatory articulation, and had more cleft-related and developmental error patterns in their speech compared to children with cleft palate and age-appropriate expressive language. Morris and Ozanne (2003) did not identify any cleft-related error patterns in the speech of children with cleft palate and age-appropriate expressive language skills. This result is contrary to what was found in the current study. Morris and Ozanne (2003) reported that the following cleft-related patterns were noted in the speech of children with cleft palate and delayed expressive language: medial consonant deletion, glottal insertion and nasal preference. These error patterns correspond to glottal errors and manner preference for nasals in SIP-CCLP Ver.3. Of the 12 children with cleft palate and expressive language scores within normal limits for their age in this study, glottal errors were identified for four children and 18 manner preference errors (for nasals) were identified for six children. These children ranged in age from 45 months to 64 months, with three children older than 3 years.

Whitehill and Chau (2004) reported that the phonetic contrast of stops versus nasals and stops versus fricatives/affricates were most often identified as being in error in their subjects with cleft palate. In the SIP-CCLP, these contrast items are included in the error type: manner preferences. For the children with cleft palate in this study, a high percentage of errors (31.6%) were in this category. The majority of errors made by children with cleft palate in this study were of the error type: place preferences (35.4%). Of note is that only one child with cleft palate was judged to have a "cleft-related" voicing error and that some children without cleft palate were judged to make "cleft-related" glottal errors. As glottal errors included all errors in which listeners did not hear a consonant in the targeted position, some of these errors may be instances of the developmental pattern in which a final consonant is deleted (e.g., "mat" heard as "ma"), while other are instances of a non-developmental pattern where initial consonants are deleted (e.g., "goat" heard as "oat"). Appendix P compares the error types tested in Whitehill and Chau (2004) and the current study in order of most to least often identified.

Whitehill and Chau (2004) included several phonetic contrasts in their intelligibility test for speakers with cleft palate that were not sampled in the SIP-CCLP (substitution of fricatives for affricates and substitution of nasals for fricatives). Listeners identified five instances of fricative for affricate substitution (e.g., "chew" identified as "zoo") for the group of speakers with cleft palate in this study. This pattern could also be classified as a three-feature error (i.e., place, manner and voicing change). Substitution of nasals for fricatives was not noted in the current study; however, for one child (CP11), substitution of fricative/nasal cluster for a fricative was noted (e.g., "zip" identified as "snip"). These errors were classified as cleft-related: other. Based on results from SIP-CCLP and Whitehill and Chau (2004), it is recommended that a phonetic contrast item be added to the closed-set response task to test the cleft-related error pattern: two-feature (place and manner) or three-feature (place, manner and voicing) substitution of fricatives for affricates. As "chew" is already elicited as part of SIP-CCLP administration, the word pair "chew- sue" or "chew – zoo" could be added to the closed-set response task to test this cleft-related manner preference error.

As well, it is recommended that a target phonetic contrast testing the cleft-related manner preference in which nasals are substituted for liquids be added to the SIP-CCLP closed-set response task. In the current study, nine instances of this error pattern were identified. To test this contrast, the contrast pair "lap – nap" could be used, as "lap" is already elicited during SIP-CCLP administration.

All of the children recruited for this study had nonsyndromic cleft palate. Both Scherer et al. (1999) and D'Antonio et al. (2001) reported differences in the percentage of consonants correct and number of glottal errors for children with and without a diagnosed syndrome. Glottal errors are classified as a cleft-related error pattern in SIP-CCLP. Data from two children with syndromic cleft palate and age-appropriate expressive language were also collected using SIP-CCLP but not included in the group comparisons. As shown in Table 12, twenty cleft-related errors were identified for these two children. Of these errors, eight were classified as glottal errors. Of all the 12 children with nonsyndromic cleft palate in this study, eight glottal errors were identified by listeners. These findings suggest that recruitment of children with syndromic cleft palate and children with cleft palate and expressive language skills below age expectations is necessary to evaluate the ability of SIP-CCLP Ver. 3 to identify the presence of other compensatory articulation patterns and cleft-related error patterns.

The majority of developmental errors identified for the children with cleft palate were in the "other" category (29.3%). This category included three instances of deaffrication (i.e. "cheep" is heard as "sheep"), 18 instances of liquid simplification and one instance of a consonant cluster identified as an affricate. Based on these results, it is recommended that future versions of SIP-CCLP include a phonetic contrast that tests deaffrication (e.g., "chew – shoe") and liquid simplification (e.g, "rail – whale"). Place preference errors accounted for the next greatest percentage of developmental errors

(22.7%). The majority of these place preference errors were substitution of alveolars for velars.

Morris and Ozanne (2003) reported that the children with cleft palate and age-appropriate language skills showed primarily the following developmental processes: cluster reduction, later stopping, liquid simplification and vocalization. In later stopping, "stops are substituted for the fricatives, /v, ð , θ/ or for affricates" (p. 470, Morris and Ozanne, 2003). Cluster reduction corresponds to developmental cluster errors and later stopping to developmental manner preference in SIP-CCLP Ver. 3. Three instances of each of these error patterns were noted for the children with cleft palate in the current study. Liquid simplification and vocalization (i.e., liquids in word final position produced as vowels or deleted) were not specifically tested in SIP-CCLP Ver.3. However, listeners identified 14 instances of liquid simplification using the "blank" response option for the 12 children with cleft palate in the current study. No instances of vocalization were noted by listeners.

"Unknown" errors in which errors not attributed to cleft palate or found during normal development were also tested in SIP-CCLP Ver. 3. The majority of "unknown" errors identified for the children with and without cleft palate were voicing errors (67.9% and 55.5%, respectively). Due to conflicting information in the literature regarding the nature of voicing errors for children with cleft palate, voicing errors where initial voiceless sounds are substituted for voiced sounds (e.g., "tear" for "dear") were included in the unexpected category (see McWilliams, Morris & Shelton, 1990; O'Gara & Logemann, 1988). Lucky (1993) found that listeners' ability to accurately identify a sound as voiced (i.e., /d/) or voiceless (e.g., /t/) increased with speaker age. Younger speakers were found to have greater variability and overlap in acoustic characteristics between these two consonants. As ten of the twelve children in each of the two groups in the current study were less than 60 months, there may be greater variability in the acoustic characteristics of their productions of voiced and voiceless sounds, leading to increased identification of these sounds as being in error. As a result, it is suggested that voicing errors remain in the unexpected category until more data from older children with and without cleft palate are collected.

Inclusion of the "blank" response option in the SIP-CCLP closed-set response task provided the opportunity for listeners to identify errors that could not be classified according to the predetermined categories described for SIP-CCLP. These unclassified errors would not have been identified if listeners had been provided with the choice of the target, two foils and "can't identify" as occurred in earlier versions of SIP-CCLP. These unclassified errors provide additional information on error patterns that may be present in the speech of children with cleft palate. Of the 15 unclassified errors, 12 involved changes to more than one feature of the target sound (place, manner, voicing). For the children without cleft palate, all but one error identified involved only one feature change. This result suggests that two or three feature changes may be unique to children with cleft palate. However, before these errors can be classified as cleft-related or developmental, further research with a larger group of children with cleft palate is necessary.

Sound analysis of error patterns.

As expected, children with cleft palate had a significantly smaller percentage of obstruents correct than the children without cleft palate. Further analysis revealed that children with cleft palate had a significantly smaller percentage of stops and fricatives correct than the children without cleft palate. However, no significant difference in percent affricates correct was found between the two groups of children. Affricates are considered a later developing sound, often not produced correctly until 7 years of age. As a result, many of the children without cleft palate may still be acquiring affricates, resulting in similar percent affricates correct between the two groups. Furthermore, there were fewer sampling opportunities for affricates, compared to stops and fricatives, in SIP-CCLP, resulting in less ability to discriminate between the two groups of children.

When examined by place of articulation, children with cleft palate had a significantly smaller percentage of alveolars correct than the children without cleft palate. This result is in keeping with expectations that children with cleft palate would have particular difficulty with alveolar sounds due to the structural abnormalities of the alveolar ridge associated with cleft palate. Children with cleft palate also had a significantly lower percent voiced and voiceless consonants correct than the children without cleft palate. Percent voiced consonants correct and percent voiceless consonants correct were similar for the children with cleft palate. This result is contrary to both McWilliams, Morris & Shelton (1990) who reported that children with cleft palate make more errors on voiceless than voiced sounds and O'Gara & Logemann (1988) who stated that these children produce voiceless sounds earlier than voiced.

Distortion analysis.

As expected, more correct and incorrect items with a distorted rating were identified for the children with cleft palate than for the children without cleft palate. These results suggest that listeners are able to identify sound distortions present in the speech of children with cleft palate and that this rating procedure is sensitive to the nonstandard errors and distortions that are observed in their speech. Interjudge agreement for "clear" and "distorted" ratings showed the same pattern as for the sound target identification, that is, lower reliability for the children with cleft palate. It might be expected that listeners would be less reliable in identifying distortions in speech that contains more sound errors and is more difficult to understand, which may be the case for the speech of the children with cleft palate. However, the lower reliability for the "clear-distortion" ratings compared to the sound identification tasks in both groups of children suggests that identification judgments can be made more reliably across listeners than distortion judgments. This is supported by Shriberg and Lof (1991) who concluded that broad transcription (phoneme identification) was much more reliable than narrow transcription (use of diacritics to indicate allophonic variations, including distortions), based on their comprehensive study of the reliability of phonetic transcription of children's speech.

In the current study, the variables "correctness" (i.e., intelligibility score) and distortion (i.e., clear/distorted rating) were examined separately. L. Hancock and M. Hodge (personal communication, April 2005) proposed a procedure to capture both "correctness" and distortion in the single variable "accuracy" that is based on a confidence rating scale for a phoneme identification task described by Yorkston, Beukelman, Bell and Strand (1999). This measure would reflect all three error types in the speech of children with cleft palate: omissions, substitutions and distortions. This description of accuracy is most similar to Shriberg, Austin, Lewis, McSweeny and Wilson's (1997) description of percent consonants correct (PCC) in which all of these patterns were identified as errors. However, in the procedure proposed by Hancock and Hodge, different weights are given to distortions, omissions and substitutions. To evaluate the validity of this variable, accuracy was determined for the 12 children with cleft palate and the 12 children without cleft palate in the current study.

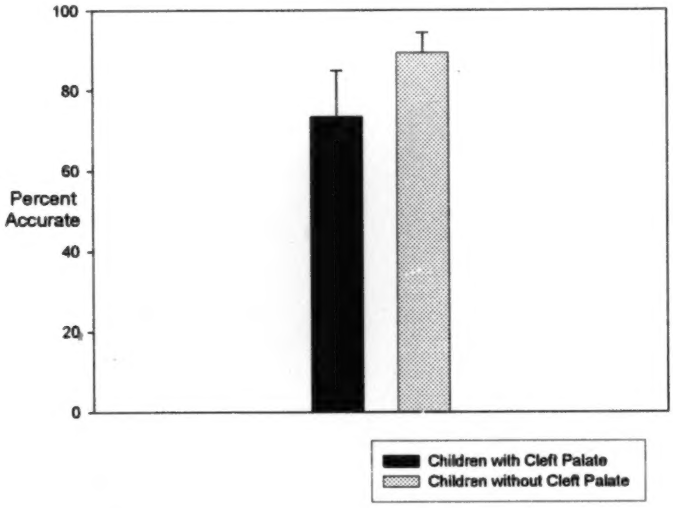
To obtain an accuracy score, listener responses in the closed-set response task were recoded such that the listener response "correct/clear" (i.e., "1") was given 2 points, the response "correct/distorted" (i.e. "2") was given one point and all incorrect responses (i.e., "3", "4", "5", "6" and "7") were given zero points. Points were then tallied for each listener. The number of points out of the total possible (i.e., number of contrast items x 2, for SIP-CCLP Ver. 3 = 194 x 2 = 388) were calculated for each listener and then the mean of the three listeners was calculated for each child and converted to a percent.

Accuracy scores for the 12 children with cleft palate (*M* = 73.4%, *SD* = 11.6) were significantly lower than the accuracy scores for the 12 age-similar children without cleft palate (*M* = 89.4%, *SD* = 5.0; t = -4.41, p< .000) with an effect size of 1.27. This effect size is greater than that for the open-set scores. These results are shown in Figure 24. Accuracy scores ranged from 58.1% to 92.8% for the children with cleft palate and from 81.9% to 96.2% for the children without cleft palate.

To determine intrajudge agreement for accuracy, repeated items in the closed-set response task were examined. Repeated items in which the listener chose the same response and distortion rating were considered to be in agreement. Calculation of intrajudge agreement for the closed-set task accuracy scores was possible for seven children with cleft palate (21 listeners) and seven children without cleft palate (21 listeners). Mean intrajudge agreement, based on the 20 repeated items, was 84.6% (*SD* =

9.8) for the 42 listeners. When examined by group, mean intrajudge agreement was 78.6% (*SD* = 10.3) for the listeners for the seven children with cleft palate and 90.7% (*SD* = 3.8) for the listeners for the seven children without cleft palate. Mean intrajudge agreement ranged from 63.3 % to 90% for the listeners for the children with cleft palate and from 88.3% to 96.7% for the listeners for the children without cleft palate.

Figure 24. Accuracy scores for age-similar children with and without cleft palate.



Note. **p<.01, one-tailed test.

Accuracy scores for the children with and without cleft palate were found to be correlated significantly with intelligibility scores from both the SIP-CCLP closed-set and open-set response tasks (r = 0.95, p < .000 for both correlations). Using a paired samples t-test, accuracy scores and intelligibility scores for both tasks were found to be significantly different (closed-set: t = 9.67, p < .000; open-set: t = -10.86, p <.000). All children had lower accuracy scores than closed-set intelligibility scores. For the children with cleft palate, this difference ranged from 5.9% to 17.3% with a mean of 11.2 (*SD* = 4.1) for the closed-set intelligibility scores. For the children without cleft palate, this difference ranged from 1.6% to 11.3% with a mean of 5.9% (*SD* = 2.8). Accuracy scores were higher than open-set intelligibility scores for all children. For the children with cleft palate, this difference ranged from 9.9% and 33.8% with a mean of 20.1 (*SD* = 6.8). For the children without cleft palate, this difference ranged from 3.8% and 26.1% with a mean of 12.7 (*SD* = 6.3). Figures 25 and 26 show the relationship between the closed-set and open-set intelligibility scores and accuracy scores.

Figure 25. Relationship between SIP-CCLP closed-set intelligibility scores and accuracy scores for children with and without cleft palate.

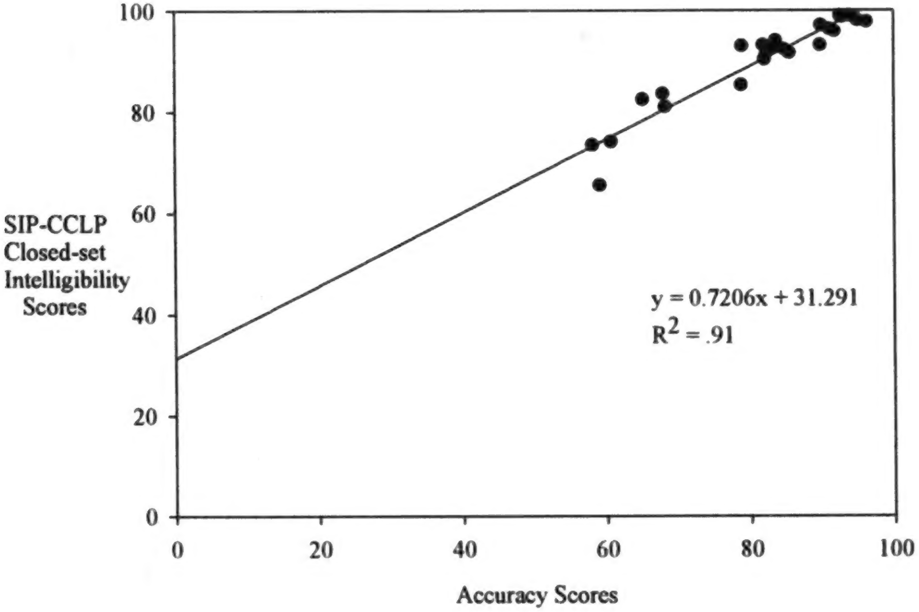
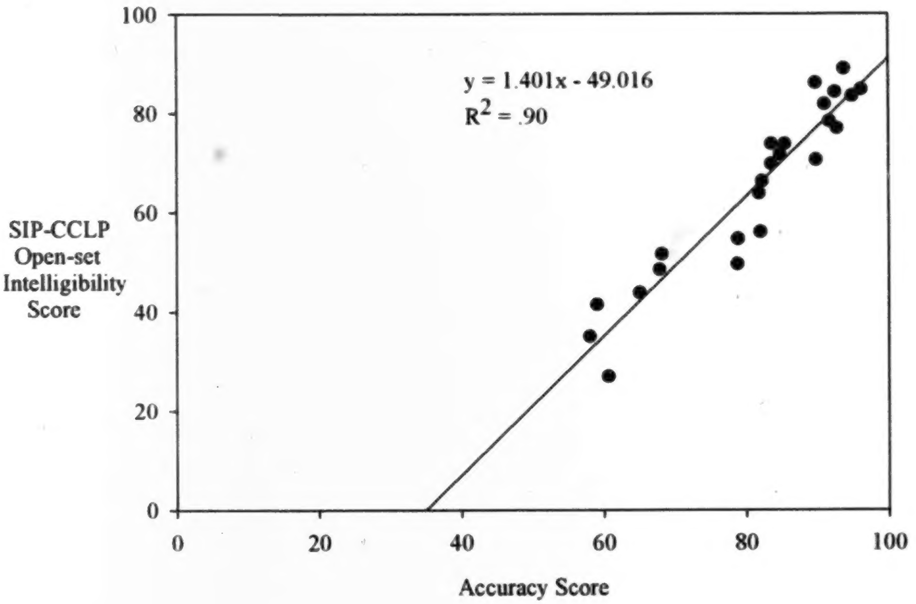
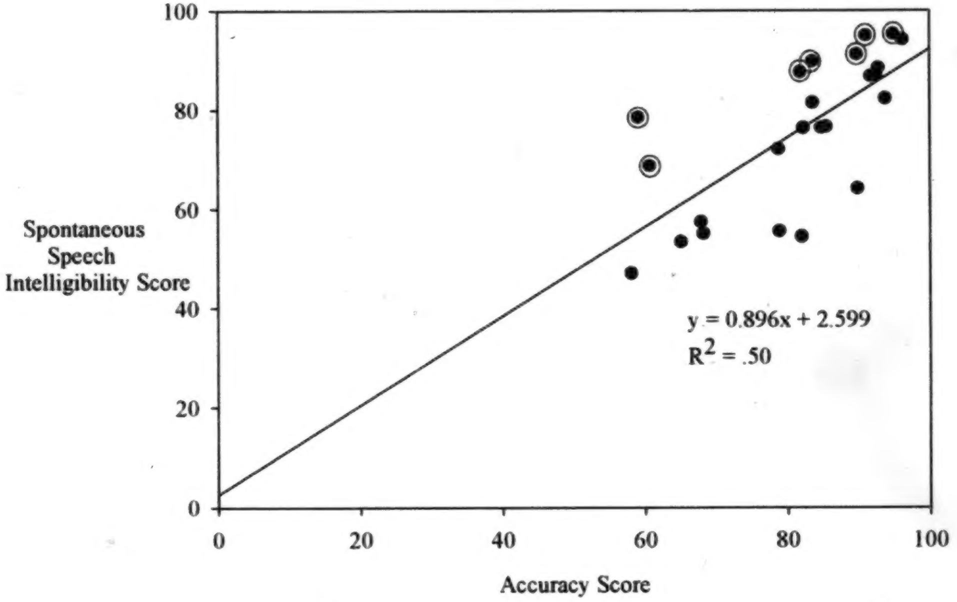


Figure 26. Relationship between SIP-CCLP open-set intelligibility scores and accuracy scores for children with and without cleft palate.



Accuracy scores for the children with and without cleft palate were found to be correlated significantly with intelligibility scores from the spontaneous speech sample (r = .71, p<.000). Using a paired samples t-test accuracy scores and spontaneous sample intelligibility scores were found to be significantly different (t = -2.67, p = .014). For the children with cleft palate, this difference ranged from 2.2% to 19.3% with a mean of 10.5% (*SD* = 6.1). For the children without cleft palate, this difference ranged from 0.3% to 27.7% with a mean of 8.7% (*SD* = 8.7). Two children with cleft palate and five children without cleft palate (circled) had lower accuracy scores than spontaneous speech intelligibility scores. Figure 27 shows the relationship between the accuracy scores and spontaneous speech open-set intelligibility scores.

Figure 27. Relationship between spontaneous speech intelligibility scores and accuracy scores for children with and without cleft palate.



Note. Circled points are those children whose spontaneous sample intelligibility scores were higher than their accuracy scores.

Conclusions

The analysis of the SIP-CCLP closed-set results revealed new information regarding the nature of errors identified in the speech of children with cleft palate. These results also provided information about the speech of children with cleft palate and typical expressive language skills that is contradictory to that reported by Morris and Ozanne (2003). As well, the analysis of the error patterns by voicing revealed no effect of voicing on frequency of error, which is contrary to results from other studies.

Despite having fewer subjects than was recommended by the power analysis for the Gotzke (2003) data, significant group differences were obtained in the expected direction for key dependent variables except number of cleft-related errors. Effect sizes were comparable to those found for Gotzke (2003) and ranged from 0.9 and 1.3. Variability in the group of children with cleft palate was high, which may be reflective of the heterogeneous profile of the speech of children with cleft palate in general. The results obtained in this study are interpreted as support for the construct validity of SIP-CCLP.

Comparison of Children without Cleft Palate by Age Group

Intelligibility scores on the SIP-CCLP open-set response task were significantly lower for youngest children (36 – 47 months) compared to the scores for the oldest children (72 – 84 months). Examination of the group means for the four age groups reveals a developmental trend in which speech intelligibility increased with age. Contrary to expectations, there was no significant difference in percent contrast items correct in the closed-set response task among the four age groups of children. However, in fitting with developmental expectations, mean percent phonetic contrast items correct increased as age increased and mean sound errors decreased. Of interest is that no child in the sample received an intelligibility score or phonetic contrast items correct score of 100%, that is, there was no ceiling effect for SIP-CCLP scores even for the children in the oldest group (72 – 84 months).

All of the children without cleft palate and all but one of the children with cleft palate recruited for the study were able to complete the task. However, several children remarked that it was too long. For some of the younger children, it was necessary to provide additional short sticker breaks to encourage them to finish. Multiple cues (verbal and visual) were needed for some of the younger children, as they often repeated the word before the frog puppet and "beep" signal occurred. Having the child repeat the word a second time seemed to be an effective strategy to ensure that the child's word production and the beep did not overlap. Children enjoyed both the animations and pictures and would often give short narratives about them. Two children commented that the picture for blue sock was actually a blue and red sock. It is recommended that this picture be recolored in future versions of SIP-CCLP.

Criterion Validity

Intelligibility

The moderately strong positive relationship between the SIP-CCLP and spontaneous speech intelligibility scores suggests that the SIP-CCLP provides a relative ranking of intelligibility scores similar to what would be obtained from these child's self-generated utterances. This finding is interpreted as support for the criterion validity of the SIP-CCLP. The finding of higher intelligibility scores for the spontaneous speech sample compared to a single word sample is similar to that reported by Gordon-Brannan (1994) who attributed this difference to the facilitating effect of context on word identification by listeners. The correlation between the two intelligibility scores is less than that reported by Gotzke (2003) (i.e., r = 0.89) and is likely related to the larger sample of children in the current study. Interjudge reliability for the listeners for the 24 children was similar to that reported by Konst et al. (2000) for their group of listeners for 20 children with and 8 children without cleft palate, aged 2.5 years (Cronbach's alpha = 0.99). When examined by group, the strength of the correlation between the SIP-CCLP and spontaneous speech intelligibility scores decreased for both groups of children. This result may be due to the small sample size and smaller range of scores within each group compared to the two groups combined. It is interesting to note that no child had low intelligibility scores for the spontaneous sample and high intelligibility scores for the SIP-CCLP open-set response task.

As SIP-CCLP is a word imitation measure, it is not intended to measure children's speech intelligibility in connected speech. The facilitating effect of context, children's use of language, the consistency of their speech error patterns and their prosody may all contribute to increased intelligibility scores for connected speech samples. With single words, as in the SIP-CCLP, many of these additional cues are lost. Furthermore, listeners had the opportunity to listen to each child's spontaneous utterances twice, whereas in the SIP-CCLP only one opportunity was given. Despite these differences, a significant relationship was found between the two samples. The open-set intelligibility scores on the SIP-CCLP and the proposed accuracy scores for the SIP-CCLP closed-set response task showed greater groups differences (larger effect size) than the open-set intelligibility scores for the SIP-CCLP and spontaneous speech sample. SIP-CCLP scores appear to have the advantages of greater time efficiency and sensitivity compared to scores from spontaneous speech samples.

Weiss (1980) provided guidelines for interpreting intelligibility scores obtained from a 100-word spontaneous sample. Weiss also stated that children between the ages of 36 and 48 months would be expected to have intelligibility scores greater than 75%. All but one of children aged 36 – 48 months in the current study had an intelligibility score in this range. By 48 months, Weiss indicated that children with typical articulation are expected to have intelligibility scores between 91% and 100%. In the current study, nine of the 18 children without cleft palate 48 months and older had intelligibility scores within this range. Despite these differences, a developmental progression in which intelligibility scores increased with increased age was found for the children without cleft palate.

Phonetic Analysis

As predicted, the mean PCC for the SIP-CCLP closed-set response task, the phonetic transcription of the SIP-CCLP words and the phonetic transcription of the 100-word spontaneous sample were not significantly different for the children with cleft palate. Mean PCC for the SIP-CCLP closed-set response task and the phonetic transcription of the spontaneous sample were not significantly different for the children without cleft palate. This result is in keeping with those found by Johnson, Weston & Bain (2004) in their comparison of PCC scores of children with speech delay on imitative sentences and spontaneous speech. This result is interpreted as support for the criterion validity of SIP-CCLP.

However, for the children without cleft palate, PCC for the phonetic transcription of the SIP-CCLP words was significantly higher when compared to the PCC for the phonetic transcription of the spontaneous sample. This difference is contrary to results reported by Drozda et al (2002) who found that PCC scores were higher in the spontaneous condition than in the imitative condition. For one child without cleft palate, PCC was higher for the phonetic transcription of the spontaneous sample than for the phonetic transcription of the SIP-CCLP words. Post-hoc analyses revealed that percent stops correct was significantly higher for the SIP-CCLP words than for the spontaneous sample, which is contrary to Morrison and Shriberg (1992), who did not find a significant difference in percent stops correct between articulation testing and spontaneous speech. However, a summary of the literature in which phonetic accuracy in citation (articulation tests) and spontaneous speech were compared (Morrison & Shriberg, 1992) found that deletion of consonants in word final position was more common in spontaneous speech. When examined by word position, percent stops correct in final position was significantly higher for the SIP-CCLP words than the spontaneous sample, supporting Morrison and Shriberg's conclusion. It is possible that for the children with typical speech, more deletion errors were noted for word final stops in conversation than in the SIP-CCLP words. A more in-depth analysis is necessary to summarize the kinds of errors made by these children on stops in word final position, to decide if this conclusion is supported.

For both the SIP-CCLP words and the spontaneous sample, interjudge agreement for consonant transcription was lower for the children with cleft palate than for the children without cleft palate. Overall transcription reliability for both speech samples for all children (12 children with cleft palate and 25 children without cleft palate) is comparable to results obtained by D'Antonio, Scherer, Miller, Kalbfleisch and Bartley (2001). D'Antonio et al. reported point-by-point interrater reliability of 82% for broad transcription (using diacritics for nasality and compensatory articulation patterns) of consonants in spontaneous speech for a group of children with and without cleft palate, suggesting that interjudge agreement obtained for the current study is acceptable.

Whitehill and Chun (2002) reported a strong correlation between PCC and intelligibility scores obtained from an imitative word sample ($r = 0.77$, $p < .001$). In the current study, a strong significant positive correlation was found between these two variables for the SIP-CCLP (r = 0.87, p <.000). When examined by group, a moderately strong significant positive correlation was found between SIP-CCLP PCC and intelligibility scores for the children with cleft palate (r = 0.78, p = .001) and a significant moderate positive correlation was found for the children without cleft palate (r = 0.69, p = .007).

A moderately strong significant positive correlation was also found between percent consonants correct and intelligibility scores for the spontaneous speech sample (r = 0.728, p<.000). For the children with cleft palate, a moderate significant positive correlation of 0.658 (p = .01) was found. For the children without cleft palate, PCC and intelligibility scores for the spontaneous sample were similar (r = 0.499, p = .049).

Recommendations for Future Revisions to SIP-CCLP

Number of Target Phonetic Contrast Items

SIP-CCLP Ver. 3 has 124 stimulus words that are recorded from a child and judged in the open-set response task. In the closed-set response task, as illustrated in Appendix B, both members of each contrast pair are judged, which results in a relatively large number of items (Version 3 = 194). An analysis was conducted to determine which items were rarely or never identified by listeners as being in error. These items are potential candidates for removal from the closed-set response task and the SIP-CCLP Ver. 3 stimulus word set. This will reduce the length of both the talkers' and listeners' tasks.

To determine if any of the target phonetic contrast items which are listed in Appendix B could be eliminated from the closed-set response task, results from a total of 135 listeners were compared. This included the listeners for the 25 children with typical speech development (75 listeners), 12 children with cleft palate (36 listeners) described in the results, as well as data for six additional children with articulation disorders (18 listeners) and 2 children with syndromic cleft palate (6 listeners). The listener responses for each target phonetic contrast were examined to determine items in which a minimum of 90% of listeners chose the target response (i.e., the same word as was said by the child, "1" or "2").

A total of 115 target phonetic contrast items that satisfied this criterion were identified: 53 items that provided opportunities for cleft-related errors, 11 items that provided opportunities for developmental errors and 51 items that provided opportunities for unknown errors. These items are in bold in Appendix B. Any items that had been identified as errors for the children with cleft palate were then removed from the pool of 115 items. This included nine cleft-related contrast items, two developmental contrast items and four unknown contrast items. The target words for the remaining 100 contrast items were then examined to determine if listeners identified any additional cleft-related error patterns using the "blank" response option for these pairs. For example, a minimum of 90% of listeners correctly identified the target word for the contrast item CED.10, but, for two children with cleft palate, listeners identified a different cleft-related error pattern (i.e., "stick" heard as "skick", "stick" heard as "snit"). Furthermore, the contrast item CED.10 is only time the target word "stick" is judged. Because of these two criteria, this phonetic contrast item was also removed from the pool of 100 items. Applying these same criteria, six additional target contrast items were eliminated from the pool (three cleft-related contrast items, two developmental contrast items, one unknown contrast item). It is suggested that the remaining 94 target phonetic contrast items be excluded from the closed-set response task, reducing the total number of target phonetic contrast items to be judged to 100 from 194. However, as a result of the analysis of the closed-set responses for this study, eight target phonetic contrast items are suggested as additions to the SIP-CCLP Ver. 4 closed-set response task, bringing the total number of target phonetic contrast items to 108. These new phonetic contrast items are listed in Table 11.

Table 11.

Suggested New Target Phonetic Contrast Items for the Next Version of SIP-CCLP (Ver. 4).

	Description of Error Pattern	Suggested Exemplar of Error Pattern
Cleft- related	Substitution of Glottal Stops for Fricatives	GEC.FA.I.XX = bash → baa
	Substitution of Glottal Fricatives for Fricatives	GEC.FH.I.XX = ship → hip
	Substitution of Glottal Fricatives for Affricates	GEC.AfH.I.XX = jail → hail
	Substitution of Liquid for Nasal	MPC.LN.I.XX = lap → nap MPC.LN.I.XX = low → no
	Substitution of Affricate for Fricative	MPC.FAf.I.XX = sue → chew
Developmental	Stopping	MPD.IB.I.XX = V → B
	Substitution of Fricative for Affricate	MPD.AfF.I.XX = chew → sue

With the reduction in number of target phonetic contrast items, the number of reliability items for the closed-set response task could also be reduced to 11 from 20 (i.e., 10% of the total number of phonetic contrast items tested). The SIP-CCLP Ver. 4 closed-set response task would then consist of four practice items, 108 target phonetic contrast items and 11 reliability items, for a total of 123 judgments. It is estimated that this reduction in the number of items will decrease the length of time it takes to complete the closed-set response task from 18 to 12 minutes. Appendix B contains the full list of target phonetic contrast items for SIP-CCLP Ver. 3. The 94 contrast items underlined and in bold are those suggested to be excluded from the next version of SIP-CCLP.

If these 94 phonetic contrast items are removed from the closed-set response task, 42 words could be removed from the SIP-CCLP Ver. 3 stimulus word set. Words in bold in Appendix A are those suggested to be excluded from the next version of SIP-CCLP. Removal of these words would reduce the number of words administered to the children to 82 and the administration time by a third. In turn, the number of listener judgments on the open-set response task would be decreased. With a reduction in the number of stimulus words, the number of reliability items for the open-set response task could also be reduced to 8 from 12. The SIP-CCLP Ver. 4 open-set response task would then consist of four practice words, 82 words and 8 reliability items, for a total of 94 judgments. It is estimated that this reduction in the number of SIP-CCLP Ver. 3 words would decrease the length of time it takes to complete the open-set response task from 15 to 12 minutes.

If these words are removed from SIP-CCLP Ver. 3, it is possible that the phonetic content of the SIP-CCLP Ver. 4 will change. The phonetic content of the SIP-CCLP Ver. 3 word set was originally designed to include a higher proportion of obstruents, as children with cleft palate make more errors on these sounds. The phonetic content of the SIP-CCLP Ver. 3 words and the targets in the closed-set response task were found to contain more obstruents when compared to the conversational speech of children in grade one (Shriberg & Kent, 1982). The changes to the phonetic content as a result of the suggested reductions in the number of SIP-CCLP Ver. 3 stimulus words and the target phonetic contrast items are shown in Table 12. The SIP-CCLP Ver. 4 82 word set contains a slightly higher proportion of obstruents and a slightly lower proportion of sonorants, which is similar to the phonetic content of the current 124 words used in SIP-CCLP Ver. 3. For stops and affricates, fricatives and nasals, the percentage of occurrence increased in the 82 word set. For the reduced number of the targets in the closed-set contrast items, the percentage of target obstruents increased compared to the original 194 target items. Reanalysis of the SIP-CCLP data for the children in the current study is necessary to determine if the suggested changes to the SIP-CCLP Ver. 3 open-set and closed-set response tasks affect their intelligibility and error scores.

Table 12.

Frequency Distribution of the Phonetic Content of SIP-CCLP Stimulus Words, the Targets in the Closed-Set Contrast Items for the Current and Proposed Versions and of Conversational Speech for Children in Grade 1.

	% Occurrence in SIP-CCLP		% Occurrence in Closed Set Response Task Judgments		% Occurrence in Conversational Speech[1]
Sound Class	124 stimulus words	82 stimulus words	194 judgments	108 judgments	
Stops and Affricates	46.46	47.4	42.43	48.5	32.65
Fricatives	18.34	19.06	26.34	31.82	25.03
Nasals	10.15	10.41	2.92	3.03	22.41
Liquids and Glides	24.99	23.12	22.93	16.67	20.16
Obstruents	64.82	66.46	68.77	80.3	57.68
Sonorants	35.14	33.53	31.22	19.7	42.57

Note. [1]Data are from Clinical Phonetics (p. 352), by L. Shriberg and R. Kent, 1982, New York: John Wiley & Sons.

Analysis and Graphing Software

In this study, data were analyzed and graphed using Microsoft Excel (Microsoft, 2000). The time invested in this has resulted in models and templates for the desired software features that will expedite programming these features into the software in the next revision.

Summary of Recommended Revisions for SIP-CCLP Ver.4

The revisions to the number of target phonetic contrasts in the closed-set response task and SIP-CCLP words are recommended, however, reanalysis of the intelligibility and error scores that result from using this subset of items is essential before these changes are instituted. Additional target phonetic contrasts should also be added to reflect the "listener generated" error types identified for the children with cleft palate and the error patterns identified by Whitehill and Chau (2004). As well, the SIP-CCLP Ver. 3 software should be revised in the next version to allow software-driven analysis of listener responses. It is expected that these changes will increase the ease with which listener responses are scored, compared and analyzed. Software changes should also allow automatic calculation of the dependent variable "accuracy". It is expected that these changes will increase the meaningfulness of the results obtained from the SIP-CCLP Ver. 4 closed-set response task.

Results for Two Children with Syndromic Cleft Palate

Data from two children with cleft palate and a diagnosed syndrome were collected but not included in the group analysis. One child (CP04) was diagnosed as having ectrodactyly-ectodermal dysplasia-clefting syndrome (EEC syndrome). This syndrome is a genetic disorder in which the most common symptoms include missing or irregular fingers and/or toes (ectrodactyly), nasolacrimal duct obstruction, cleft lip and palate and urogenital abnormalities (United States National Library of Medicine , 1999). Mental retardation, conductive hearing loss and abnormalities of the ears and face may also be associated. Child CP04 had a repaired bilateral cleft lip and palate, normal hearing in both ears (reported by audiologist) and receptive and expressive language within normal limits, based on results from Preschool Language Scale Version 4, PLS-4, (Zimmerman et al, 2002).

The other child, CP12, was diagnosed as having Stickler syndrome. Stickler syndrome is a genetic disorder with variable expressivity, which classically presents as Pierre Robin sequence with cleft palate, early onset osteoarthritis and nearsightedness (Kummer, 2001). Characteristic facial features include micrognathia (i.e., small lower jaw), a flat facial profile, and midface hypoplasia (underdevelopment). CP12 had a repaired cleft of the soft palate, mild-to-moderate hearing loss in both ears and receptive and expressive language within normal limits. Results for the SIP-CCLP Ver. 3 and spontaneous sample for these two children and the twelve children with nonsyndromic cleft palate in the current study are shown in Table 13.

Table 13.

SIP-CCLP Ver. 3 and Spontaneous Sample Scores for Two Children with Syndromic Cleft Palate and Twelve Children with Nonsyndromic Cleft Palate.

		Children with nonsyndromic cleft palate (n = 12)	CP04 (41 months)	CP12 (60 months)
SIP-CCLP Open-set Response Task	Intelligibility Score – SIP-CCLP open-set	53.3[1] (*SD* = 15.9)	23.9**	29.3**
SIP-CCLP Closed-set Response Task	Intelligibility Score	84.5 (*SD* = 9.9)	73.9*	68.0**
	Percent phonetic contrast items correct	86.5 (*SD* = 9.8)	77.3	69.6**
	Number of cleft-related errors	6.7 (*SD* = 7.6) Range: 0 – 20 Total Number: 79	3	17
	Percent obstruents correct	85.4 (*SD* = 11.5)	73.8*	61.0***
	Percent of correct items given a distorted rating	22.9 (*SD* = 8.2)	29.7	35.1*
	Percentage of incorrect items given a distorted rating	6.55 (*SD* = 4.7)	13.8**	19.8***
	Accuracy Score	73.4 (*SD* = 11.6)	58.9*	50.7**
Spontaneous Sample	Intelligibility Score	67.5 (*SD* = 13.3)	71.8	76.4

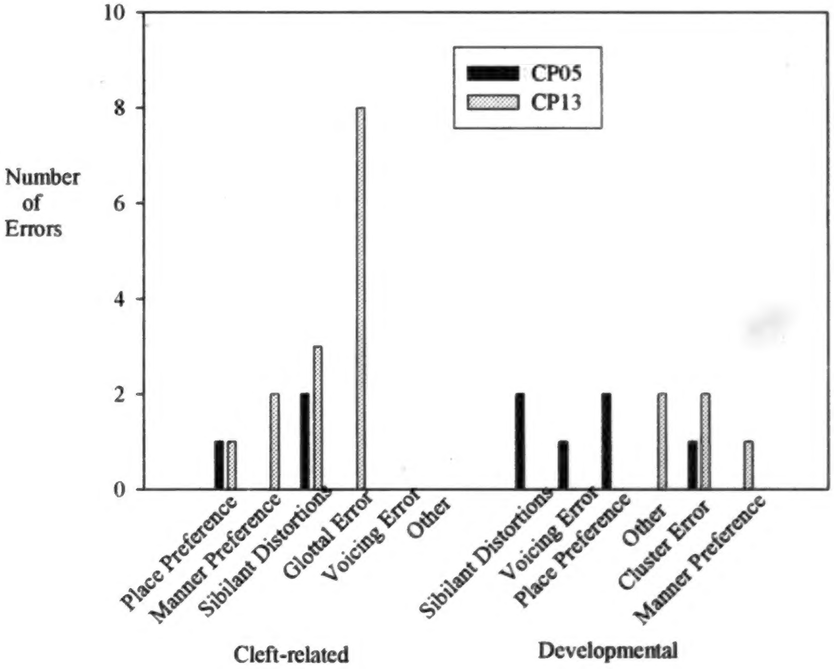
Note. [1]Means are reported, unless otherwise indicated. Stars indicate number of standard deviations from the mean child's score is from the children with nonsyndromic cleft palate (* more than 1 standard deviation from the mean, **more than 1.5 standard deviations, *** more than 2 standard deviations).

On the SIP-CCLP Ver. 3, these children had lower intelligibility scores on the open-set and closed-set response tasks, lower percent phonetic contrast items correct, lower percentage obstruents correct and greater percentage of correct and incorrect items given a distorted rating compared with the group of children with nonsyndromic cleft palate. It is interesting to note that the intelligibility scores for the spontaneous sample were within one standard deviation of the mean for the children with nonsyndromic cleft palate, whereas the SIP-CCLP intelligibility scores for these children was more than 1.5 standard deviations from the mean. The difference between SIP-CCLP Ver. 3 open-set and spontaneous sample intelligibility scores for the two children was 47.1 (CP04) and 47.9 (CP12). When results for these two children are included in the calculation of the regression equation, a significant relationship between SIP-CCLP open set and spontaneous sample intelligibility scores was not found (r = 0.40, p = .078). As the SIP-CCLP Ver. 3 has a greater percentage of obstruents than would be found in spontaneous speech (Table 12), there is increased emphasis on the sounds that children with cleft palate often have difficulty with. This may result in lower SIP-CCLP Ver. 3 open-set intelligibility scores than would be expected based on spontaneous sample intelligibility scores. As well, other factors such as prosody, voice, resonance, consistency of error patterns and language use can affect intelligibility scores.

Error profiles for these two children are shown in Figure 28. For child CP04, unlike the children in the current study, two instances of the error category “Unknown manner preference error”, in which nasals are identified as oral stops, were identified. It would be expected the child CP04 would have hypernasal speech related to cleft palate. However, this pattern of nasal to oral suggests that this child’s voice is also hyponasal, which may be related to the abnormalities of the oropharynx and nasopharynx associated with EEC.

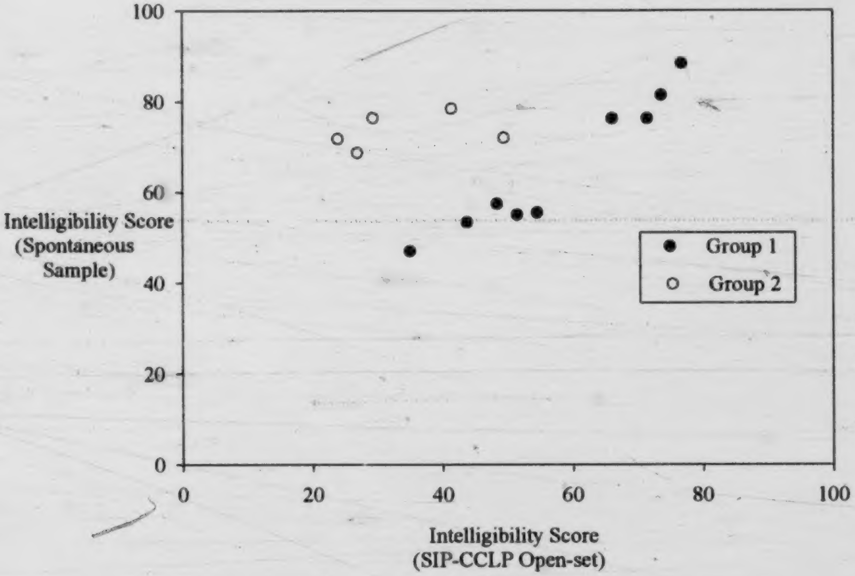
For child CP12, the majority of errors identified were in the category "Cleft-related: glottal error". These errors included five instances in which fricatives (i.e., /s, ʃ/) were identified as the glottal fricative (/h/) and one instance where an affricate (i.e., /ʤ/) was identified as /h/. This error pattern was not identified for the children with nonsyndromic cleft palate. Four errors were not classifiable: one instance of a cluster (/bl/) identified as a velar (/k/), one instance of a velar (/g/) identified as a labiodental fricative (/f/) and two instances of a nasal for fricative substitution (i.e., "flow" identified as "mow" and "nut fell" identified as "nut nell"). Whitehill and Chau (2004) found that the nasal for fricative error pattern had the highest mean error proportion for the cleft palate speakers in their study.

Figure 28. Contrast target error profile based on results from SIP-CCLP Ver. 3 for CP04 and CP12.



Visual examination of the scatter plot shown in Figure 29 suggests that there may be two groups of children with cleft palate with different relationships between intelligibility scores from the two sampling conditions. Group 1 includes nine children for whom the difference between intelligibility scores was less than 12%. Group 2 includes five children, including the two children with syndromic cleft palate, for whom the difference between intelligibility scores is greater than 20%. Examination of data for the children in Group 2 did not reveal any other commonalities among these five children (e.g., different types of cleft, not the children with the lowest intelligibility scores or highest number of errors or percent correct/distorted). Recruitment of additional children with cleft palate is necessary to determine the validity of these two subgroups of children.

Figure 29. Relationship between SIP-CCLP open-set and spontaneous sample intelligibility for all children with cleft palate (including children with a diagnosed syndrome) divided into Group1 (< 12% difference between intelligibility scores) and Group 2 (>20% difference between intelligibility scores).



The information collected from these two children provides corroborative evidence for the construct validity of the SIP-CCLP. Their scores were in many cases more than one standard deviation of the mean than the scores for the children with nonsyndromic cleft palate. Results for these two children with syndromic cleft palate provide additional information about the nature and pattern of errors that might be identified in the speech of children with cleft palate.

Future Research

The results from this relatively small sample of children suggest that the SIP-CCLP shows promise as a useful tool for measuring speech intelligibility, quantifying speech sound distortions and identifying which phonetic contrast errors may be contributing to reduced intelligibility in children with cleft palate. It also allows examination of the relationship between distortion and intelligibility through calculation of "accuracy scores". Although the SIP-CCLP is not substitute for an intelligibility score based on spontaneous speech, it does provide reliable and valid measures of intelligibility that are more sensitive and less time-consuming to obtain.

In this study, the children with cleft palate were heterogeneous with respect to the type of cleft (e.g., bilateral cleft lip and palate, submucous, cleft palate only). Whitehill and Chau (2004) suggested that an intelligibility test that was designed to be sensitive to the error patterns found in the speech of children with cleft palate may be particularly sensitive to differences among groups of children with different types of clefts. Increased recruitment of children with cleft palate would allow an evaluation of effect of cleft type on intelligibility.

Replication of this study with children with cleft palate and expressive language delays and children with cleft palate and a diagnosed syndrome would be valuable to assess the differences in speech error patterns in these groups of children compared to children with nonsyndromic cleft palate and age-appropriate expressive language. Furthermore, recruitment of children with these characteristics will allow further evaluation of the revised target contrast error patterns tested in SIP-CCLP Ver. 3. None of the children in the current study produced any compensatory articulations other than glottal stops. Inclusion of children with a wider range of characteristics may allow assessment of what listeners hear when children produce compensatory articulations and allow a more comprehensive evaluation the effectiveness of the "blank" response option as a way to identify nonstandard substitution errors. As the population of children with cleft palate is heterogeneous, the inclusion of different groups of children would help ensure that the SIP-CCLP would be valid and reliable tool for all children with cleft palate.

In this study, results were evaluated on a group basis (i.e., children with and without cleft palate). However, individual children with similar intelligibility scores may have had different phonetic contrasts in error. In fact, Whitehill and Chau (2004) found that children with similar intelligibility scores had different phonetic contrast error profiles. Kent et al. (1989) suggested that one of the values of having an intelligibility tool that specifically tests the error patterns of the population of interest is that it promotes more focused treatment. The value of the SIP-CCLP closed-set response task as a way to identify possible targets for treatment is a next important step in evaluating this tool.

Conclusions

This study examined the reliability and validity of the Speech Intelligibility Probe for Children with Cleft Palate Version 3 (SIP-CCLP Ver. 3) by comparing SIP-CCLP scores for children with and without cleft palate. The children with cleft palate had lower intelligibility scores, lower percentage of obstruents incorrect and a greater percentage of correct and incorrect items given a distorted rating compared to the children without cleft palate. An age comparison of children's scores on the SIP-CCLP also revealed the predicted developmental trend where the youngest age group had lower percent correct scores than the older age groups. In support of the criterion validity of the SIP-CCLP, a significant positive correlation was found between intelligibility scores from the SIP-CCLP and the spontaneous sample. However, the specific relationship and amount of difference between intelligibility scores for these two samples varied between groups and for individuals in each group. As well, percent consonants correct was similar between phonetic transcription of the speech sample and the SIP-CCLP productions and between an error analysis of the closed-set results and phonetic transcription of the speech sample. In addition, the SIP-CCLP was found to have acceptable interjudge and intrajudge reliability. The findings of this study provide support for the validity and reliability of the SIP-CCLP as a measure of speech intelligibility for children with cleft palate. They also helped to identify recommendations to increase the sensitivity and efficiency of the SIP-CCLP Ver. 3 software for the next stage of software development (Ver. 4) and evaluation.

REFERENCES

Adobe Systems Incorporated. (2004). Adobe Audition 1.5. San Jose, CA; Adobe Systems Incorporated.

American Speech and Hearing Association (ASHA). (1985). Guidelines for identification audiometry. American Speech and Hearing Association Journal, 27, 409-453.

Bailey, C., Anderson, P., Fabbri, J., Kinder, J., Larson, G., Hoffman, J., Owen, B., Rys, C., Schneider, D., & Xue, S. (1989). Adobe Photoshop Business Edition 1.0. United States: Adobe Systems Inc.

Blishen, B.R., Carroll, W.K., & Moore, C. (1987). The 1981 Socioeconomic Index for Occupations in Canada. Canadian Review of Sociology and Anthropology, 24, 465-487.

Chapman, K.L. (1993). Phonologic processes in children with cleft palate. Cleft Palate-Craniofacial Journal, 30(1), 64-72.

Chin, S.B., Finnegan, K.R., & Chung, B.A. (2001). Relationships among types of speech intelligibility in pediatric users of cochlear implants. Journal of Communication Disorders, 34, 187-205.

Cohen, J., Welkowitz, J., & Ewen, R.B. (1991). Introductory Statistics for the Behavioral Sciences, 4th Edition. New York, Harcourt Brace Jovanovich.

Conley, S.F., Gosain, A.K., Marks, S.M. & Larson, D.L. (1997). Identification and assessment of velopharyngeal inadequacy. American Journal of Otolaryngology, 18(1), 38-46.

Connolly, S. (2001). A Phonetic Contrast Approach to Assessing Intelligibility in Children with Cleft Palate. Unpublished manuscript, University of Alberta, Edmonton, Canada.

D'Antonio, L.L., Scherer, N.J., Miller, L.L., Kalbfleisch, J.H., & Bartley, J.A. (2001). Analysis of speech characteristics in children with velocardiofacial syndrome (VCFS) and children with phenotypic overlap without VCFS. Cleft Palate-Craniofacial Journal, 39(5), 455 – 467.

Dalston, R.M. (1990). Communication skills of children with cleft lip and palate: a status report. In Bardach, J., Morris, H.L., eds. Multidisciplinary Management of Cleft Lip and Palate. Philadelphia: WB Saunders, 746 - 749.

Drozda, A., Gustafson, B., Roberts, H., Hodge, M., & Wellman, L. (2002, April). Effects of sampling condition on measures of speech function in children. Poster presented at the Canadian Association of Speech-Language Pathologists and Audiologists (CASLPA) Conference, Victoria, British Columbia, Canada.

Dworkin, J. & Culatta, R. (1996). Dworkin-Culatta Oral Mechanism Exam – Treatment (D-COME-T). Nicholasville, KY: Edgewood Press.

Feltz, C., McClure, K., & O'Hare, J. ([illegible]002). Speech Intelligibility Probe for Children with Cleft Palate (SIP-CCLP): A Preliminary Assessment of Validity and Reliability. Unpublished manuscript, University of Alberta, Edmonton, Canada.

Fluharty, N.B. (2001). Fluharty Preschool Speech and Language Screening Test (Fluharty-2). Austin, TX: Pro-Ed Inc.

Golding-Kushner, K.J. (1995). Treatment of articulation and resonance disorders associated with cleft palate and VPI. In: Shprintzen, R.J., Bardach, J., eds. Cleft Palate Speech Management: A Multidisciplinary Approach. St. Louis: Mosby, 327-351.

Goldman, R., & Fristoe, M. (2000). Goldman Fristoe Test of Articulation. Circle Pines, MN: American Guidance Service.

Gooch, J.L., Hardin-Jones, M., Chapman, K.L., Trost-Cardamone, J.E., Sussman, J. (2001). Reliability of listeners transcriptions of compensatory articulations. Cleft Palate-Craniofacial Journal, 38(1), 59 – 67.

Gordon-Brannan, M. (1994). Assessing intelligibility: Children's expressive phonologies. Topics in Language Disorders, 14(2), 17-25.

Gordon-Brannan, M. & Hodson, B.W. (2000). Intelligibility/Severity measurements of prekindergarten children's speech. American Journal of Speech-Language Pathology, 9, 141-150.

Gotzke, C. (2003). Continuing Assessment of the Validity and Reliability of the Speech Intelligibility Probe for Children with Cleft Palate Version 2 (SIP-CCLP Ver. 2). Unpublished manuscript, University of Alberta, Edmonton, Canada.

Gotzke, C. & Hodge, M. (2003). Technical Report #1: Analysis of overall results and closed-set errors on SIP-CCLP Ver. 1 for two children with cleft palate. Unpublished manuscript.

Harding, A. & Grunwell, P. (1996). Characteristics of cleft palate speech. European Journal of Disorders of Communication, 31, 331-357.

Hirschberg, J. (1986). Velopharyngeal insufficiency. Folia Phoniatrica, 38, 221-276.

Hodge, M. (1996). Measuring the speech intelligibility of young children with dysarthria. Paper presented at the 8th Biennial Conference on Motor Speech, Ameila Island, Florida.

Johnson, C., Weston, A. & Bain, B. (2004). An objective and time-efficient method for determining severity of childhood speech delay. American Journal of Speech-Language Pathology, 13(1), 55-65.

Kent, R.D., Miolo, G., & Bloedel, S. (1994). The intelligibility of children's speech: A review of evaluation procedures. American Journal of Speech Language Pathology, 3(2), 81-95.

Kent, R.D., Weismer, G., Kent, J.F., & Rosenbek, J.C. (1989). Toward phonetic intelligibility testing in dysarthria. Journal of Speech and Hearing Disorders, 54, 482-499.

Keuning, K.H., Wieneke, G.H., Van Wijngaarden, H.A. & Dejonckere, P.H. (2002). The correlation between nasalance and a differentiated perceptual rating of speech in Dutch patients with velopharyngeal insufficiency. Cleft Palate-Craniofacial Journal, 39(3), 277-284.

Khan, L. & Lewis, N. (1986). Khan-Lewis Phonological Analysis. Circle Pines, MN: American Guidance Service.

Kirk, R. (1968). Experimental Design: Procedures for the Behavioral Sciences, Belmont, CA: Brooks/Cole Publishing Company.

Konst, E.M., Weersink-Braks, H., Rietveld, T., & Peters, H. (2000). An intelligibility assessment of toddlers with cleft lip and palate who received and did not receive presurgical infant orthopedic treatment. Journal of Communication Disorders, 33, 483-501.

Kummer, A.W. (2001). Cleft Palate and Craniofacial Anomalies: Effects on Speech and Resonance. Canada: Singular.

Lucky, K.W. (1993). Perceptual and Acoustic Analysis of Word-Initial Voicing Contrasts Across Speaker Age. Unpublished manuscript, University of Alberta, Edmonton, Canada.

Macromedia. (2001). Authorware 6.0. San Francisco, CA: Macromedia, Inc.

Maegawa, J., Sells, R.K. & David, D.J. (1998). Speech changes after maxillary advancement in 40 cleft lip and palate patients. Journal of Craniofacial Surgery, 9(2), 177-182.

McWilliams, B.J. (1954). Some factors in the intelligibility of cleft-palate speech. Journal of Speech and Hearing Research, 19, 524-527.

McWilliams, B.J., Morris, H.L. & Shelton, R.L. (1990). Disorders of phonation and resonance. In: McWilliams, B.J., Morris, H.L. & R.L. Shelton (Eds.). Cleft Palate Speech, pp. 247-268. Philadelphia: B.C. Decker.

Metz, D.E., Samar, V.J., Schiavetti, N., Sitler, R.W., & Whitehead, R.L. (1985). Acoustic dimensions of hearing-impaired speakers' intelligibility. Journal of Speech and Hearing Research, 28, 345-355.

Microsoft PowerPoint. (2000). [Computer program]. Seattle, WA: Microsoft Corporation.

Microsoft Excel. (2000). [Computer program]. Seattle, WA: Microsoft Corporation.

Milenkovic, P., & Read, C. (1992). CSpeech Version 4.0, Laboratory Version. Madison, WI: University of Wisconsin-Madison.

Morris, H. & Ozanne, A. (2003). Phonetic, phonological and language skills of children with a cleft palate. Cleft Palate-Craniofacial Journal, 40(5), 460-470.

Morrison, J.A., & Shriberg, L. (1992). Articulation testing versus conversational speech testing. Journal of Speech and Hearing Research, 35, 259-273.

O'Gara, M.M. & Logemann, J.A. (1988). Phonetic analysis of the development of babies with cleft palate. Cleft Palate Journal, 25, 122-134.

Pannbaker, M. (1975). Oral language skills of adult cleft palate speakers. Cleft Palate Journal, 12(1), 95-106.

Persson, C., Elander, A., Lohmander-Agerskov, A. & Soderpalm, E. (2000). Speech outcomes in isolated cleft palate: Impact of cleft extent and additional malformations. Cleft Palate-Craniofacial Journal, 37(4): 397-408.

Peterson-Falzone, S.J. (1985). Velopharyngeal inadequacy in the absence of overt cleft palate. Journal of Craniofacial Genetics and Developmental Biology Supplement, 1, 97-124.

Peterson-Falzone, S.J. (1990). A cross-sectional analysis of speech results following palatal closure. In: Bardach, J, Morris, HL (Ed.) Multidisciplinary Management of Cleft Lip and Palate. Philadelphia: WB Saunders.

Rvachew, S., Slawinski, E., Williams, M., & Green, C. (1999). The impact of early onset otitis media on babbling and early language development. Journal of the Acoustical Society of America, 105(1), 467-475.

Scherer, N.J., D'Antonio, L.L. & Kalbfleisch, J.H. (1999). Early speech and language development in children with velocardiofacial syndrome. American Jouranl of Medical Genetics (Neuropsychiatric Genetics), 88, 714 -723.

Schiavetti, N. (1992). Scaling procedures for the measurement of speech intelligibility. In: R.D. Kent (Ed.), Intelligibility in Speech Disorders: Theory, Measurement and Management (pp. 119-155). Amsterdam: John Benjamins.

Schiavetti, N., & Metz, D. (2002). Evaluating Research in Communicative Disorders (4th Edition). Boston, MA: Allyn and Bacon.

Shriberg, L. (1986). Programs to Examine Phonetic and Phonological Evaluation Records (P.E.P.P.E.R.) Version 4.0. Madison, WI: University of Wisconsin-Madison.

Shriberg, L.D., Austin, D., Lewis, B.A., McSweeny, J.L. & Wilson, D.L. (1997). The percent of consonants correct (PCC) matrix: Extensions and reliability data. Journal of Speech, Language and Hearing Research, 40(4), 708-722.

Shriberg, L.D. & Kent, R. (1982). Clinical Phonetics. New York: John Wiley & Sons.

Shriberg. L.D. & Lof, G.L. (1991). Reliability studies in broad and narrow phonetic transcription. Clinical Linguistics & Phonetics, 5(3), 225-279.

Shrout, P.E., & Fleiss, J.L. (1979). Intraclass correlations: Uses in assessing rater reliability. Psychological Bulletin, 86(2), 420-428.

Stimley, M.A., & Hambrecht, G. (1999). Comparisons of children's single-word articulation proficiency, single-word speech intelligibility and conversational speech intelligibility. Journal of Speech-Language Pathology and Audiology, 23(1), 19-23.

Subtelny, J.D., Van Hattum, R.J., & Myers, B.B. (1972). Ratings and measures of cleft palate speech. Cleft Palate Journal, 9, 18-27.

Trost, J.E. (1981). Articulatory additions to the classical description of the speech of persons with cleft palate. Cleft Palate Journal, 18(3), 193-203.

Trost-Cardamone, J.E. (1990). Speech in the first year of life: A perspective on early acquisition. In: Kernahan, D.E. & S.W. Rosenstein (Eds.). Cleft Lip and Palate: A System of Management, pp. 91-103. Baltimore: Williams & Wilkins.

United States National Library of Medicine (1999). Retrieved May 19, 2005 from http://www.nlm.nih.gov/mesh/jablonski/syndromes/syndrome278.html

Van Demark. D.R., Morris, H.L. & Van De Haar, C. (1979). Patterns of articulation abilities in speakers with cleft palate. Cleft Palate Journal, 18, 193-202.

Vogt, W.P. (1993). Dictionary of Statistics and Methodology. Newbury Park, CA: Sage Publications.

Weiss, C. (1980). Weiss Comprehensive Test of Articulation. Hingham, Massachusetts: Teaching Resources Corporation.

Weston, A.D., & Shriberg, L.D. (1992). Contextual and linguistic correlates of intelligibility in children with developmental phonological disorders. Journal of Speech and Hearing Research, 35, 1316-1332.

Whitehill, T. (2002). Assessing intelligibility in speakers with cleft palate: A critical review of the literature. Cleft Palate-Craniofacial Journal, 39(1), 50-58.

Whitehill T.L. & Chau C. (2004). Single-word intelligibility in speakers with repaired cleft palate. Clinical Linguistics & Phonetics. 18, 341-355.

Whitehill, T. & Chun, J.C. (2002). Intelligibility and acceptability of speakers with cleft palate. In: Windsor, F., Kelly, M.L. & N. Hewlett, eds. Investigations in Clinical Phonetics and Linguistics. Mahwah: Lawrence Erlbaum Associates, 405-415.

Witzel, M.A. (1995). Communicative impairment associated with clefting. In: Shprintzen, R.J., Bardach, J., eds. Cleft Palate Speech Management: A Multidisciplinary Approach. St. Louis: Mosby, 137-166.

Wyatt, R., Sell, D., Russell, J., Harding, A., Harland, K., & Albery, E. (1996). Cleft palate speech dissected: A review of current knowledge and analysis. British Journal of Plastic Surgery, 49(3), 143-149.

Yorkston, K.M., & Beukelman, D.R. (1978). A comparison of techniques for measuring intelligibility of dysarthric speech. Journal of Communication Disorders, 11, 499-512.

Yorkston, K.M., & Beukelman, D.R. (1980). A clinician-judged technique for quantifying dysarthric speech based on single-word intelligibility. Journal of Communication Disorders, 13, 15-31.

Yorkston, K.M., & Beukelman, D.R. (1981). Assessment of Intelligibility of Dysarthric Speech. Austin, TX: Pro-Ed, Inc.

Yorkston, K.M., Beukelman, D.R., Strand, E.A. & Bell, K.R. (1999). Management of Motor Speech Disorders in Children and Adults, Second Edition. Austin, TX: Pro-Ed, Inc.

Zimmerman, I.L., Steiner, V.G., & Pond, R.E. (2002). Preschool Language Scale Fourth Edition (PLS-4). Texas: The Psychological Corporation.

APPENDIX A
SIP-CCLP Ver. 2 Stimulus Words

Practice Words (in order of presentation)

1. jet
2. read
3. shoelace
4. flower house

Test Items (in alphabetical order)

1. **ape**
2. bale
3. **base**
4. bash
5. **bat**
6. bath
7. B
8. **block**
9. **blue A**
10. **blue J**
11. blue lock
12. **blue sock**
13. bud
14. bug
15. cake
16. cap
17. cape
18. cat
19. chalk
20. cheep
21. chew
22. **clap**
23. **coat**
24. D
25. daisy
26. day C
27. davie
28. deer
29. down
30. **E**
31. face
32. fail
33. **fat**
34. **fell**
35. **fire cat**
36. **fire hat**
37. flow
38. **G**
39. **go**
40. goat
41. gown
42. grow
43. **hail**
44. **hat**
45. jail
46. jeep
47. **K**
48. kid B
49. kid E
50. **king**
51. kitty
52. knee
53. lace
54. **lake**
55. **lamb**
56. lap
57. **link**
58. lip
59. lock
60. log
61. long
62. low
63. mad
64. **man**
65. mat
66. money
67. mouse
68. mouth
69. **mucky**
70. muddy
71. mug E
72. nut fell
73. nut sell
74. nut shell
75. **O**
76. **P**
77. **pail**
78. **peach**
79. race
80. rail
81. **rat**
82. ring
83. rip
84. **road**
85. robe
86. rock
87. rope
88. **safe**
89. sail
90. **sat**
91. sell
92. shell
93. ship
94. **shoe**
95. **sink**
96. sip
97. **sip up**
98. sit up
99. slip
100. stew
101. stick
102. string
103. strip
104. tail
105. tape
106. tear
107. think
108. tick
109. toe
110. trail
111. trip
112. two
113. **U**
114. V
115. veil
116. **walk**
117. **well**
118. **whale**
119. **wing**
120. year
121. **yell**
122. zee
123. zip
124. zoo

APPENDIX B

SIP-CCLP Ver. 3 Target Phonetic Contrast Pairs (Closed-set response task)

(Note. Contrast items in bold are those in which 90% of listeners correctly identified the target. Contrast items underlined and in bold are those suggested to be excluded from next version of SIP-CCLP)

I. Cleft-Related Errors

A. Sibilant Errors

Palatalization

SEC.AP.I.1 = sip → ship
SEC.AP.I.2 = sell → shell
SEC.AP.M.3 = nut sell → nut shell

Weakening

SEC.AL.I.4 = sat → fat
SEC.AL.I.5 = sail → fail
SEC.AL.I.6 = zee → V
SEC.AL.M.7 = nut sell → nut fell
SEC.AL.M.8 = daisy → davie

B. Voicing Errors

Word Final Voicing of Voiceless Stops

VEC.S.UVo.F.7 = rope → robe
VEC.S.UVo.F.8 = mat → mad
VEC.S.UVo.F.9 = lock → log

C. Glottal Errors

Substitution of Glottal Stops for Oral Stops

GEC.Bʔ.I.1= B → E
GEC.Bʔ.M.2 = kid B → kid E
GEC.Bʔ.F.3 = jeep → G
GEC.Aʔ.I.4 = D → E
GEC.Aʔ.I.5 = toe → O
GEC.Aʔ.F.6 = goat → go
GEC.Vʔ.I.7 = cape → ape
GEC.Vʔ.F.8 = cake → K

Substitution of Glottal Stops for Affricates

GEC.Afʔ.I.9 = G → E
GEC.Afʔ.M.10 = blue J → blue A
GEC.Afʔ.F.11 = peach → P

Substitution of Glottal Fricatives for Oral Stops

GEC.BH.I.12= pail → hail
GEC.AH.I.13= tail→ hail
GEC.VH.I.14= cat → hat
GEC.VH.M.15= fire cat → fire hat

D. Place Preference

Substitution of Velars for Alveolars

PPC.AV.I.1 = down → gown
PPC.AV.I.2 = tape → cape
PPC.AV.M.3 = muddy → mug E
PPC.AV.F.4 = bud → bug

Substitution of Bilabials for Alveolars

PPC.AB.I.5 = tail → pail
PPC.AB.I.6 = D → B
PPC.AB.M.7 = sit up → sip up
PPC.AB.F.8 = cat → cap
PPC.AB.F.9 = road → robe

E. Manner Preference

Substitution of Sonorants for Stops

MPC.SGl.I.1 = pail → whale
MPC.SGl.I.2 = bail → whale
MPC.SGl.I.3 = two → U
MPC.SGl.I.4 = deer → year
MPC.SGl.I.5 = king → wing
MPC.SLi.I.6 = base → lace
MPC.SLi.I.7 = toe → low
MPC.SLi.I.8 = tail → rail
MPC.SLi.I.9 = cake → lake

Substitution of Sonorants for Fricatives

MPC.FGl.I.10 = fail → whale
MPC.FGl.I.11 = fell → yell
MPC.FGl.I.12 = zoo → U
MPC.FGl.I.13 = shell → well
MPC.FGl.I.14 = shoe → U
MPC.FGl.I.15 = shell → yell
MPC.FLi.I.16 = face → lace
MPC.FLi.I.17 = fat → rat
MPC.FLi.I.18 = face → race
MPC.FLi.I.19 = veil → rail
MPC.FLi.I.20 = sink → link
MPC.FLi.I.21 = ship → rip
MPC.FLi.M.22 = blue sock → blue lock

MPC.FLi.F.23 = safe → sail
MPC.FLi.F.24 = base → bale

Substitution of Sonorants for Affricates
MPC.AfGl.I.25= chalk → lock
MPC.AfLi.I.26= chew → U
MPC.AfLi.I.27= chalk → rock
MPC.AfLi.I.28= chalk → walk
MPC.AfLi.I.29= jail → rail

Substitution of Nasals for Oral Stops
MPC.SN.I.30 = bat → mat
MPC.SN.I.31 = D → knee
MPC.SN.M.32 = muddy → money
MPC.SN.F.33 = lap → lamb
MPC.SN.F.34 = mad → man
MPC.SN.F.35 = log → long

II. Developmental Errors
A. Sibilant Errors
Fronting
SED.PA.I.1 = ship → sip
SED.PA.I.2 = shell → sell
SED.PA.M.3 = nut shell → nut sell

Fronting
SED.AI.I.9 = sink → think
SED.AI.F.10 = mouse → mouth
SED.PI.F.11 = bash → bath

B. Voicing Errors
Substitution of Voiced Stop for Voiceless
VED.S.UVo.I.1 = P → B
VED.S.UVo.I.2 = pail → bale
VED.S.UVo.I.3 = tear → deer
VED.S.UVo.I.4 = coat → goat
VED.S.UVo.M.5 = kitty → kid E
VED.S.UVo.M.6 = mucky → mug E

Substitution of Voiced Fricative for Voiceless
VED.F.UVo.I.10 = fail → veil
VED.F.UVo.I.11 = sip → zip
VED.F.UVo.M.12 = day C → daisy

Substitution of Voiced Affricate for Voiceless
VED.Af.UVo.I.13 = cheep → jeep

D. Place Preference
Fronting
PPD.VA.I.1 = gown → down
PPD.VA.I.2 = cape → tape
PPD.VA.M.3 = mug E → muddy
PPD.VA.F.4 = bug → bud

C. Manner Preference
Stopping
MPD.FS.I.36 = sail → tail
MPD.FS.I.37= zee → D
MPD.AfS.I.38 = chew → two

E. Cluster Error
Cluster Reduction
CED.1 = block → lock
CED.2 = flow → low
CED.3 = slip → lip
CED.4 = trail → rail
CED.5 = trip → rip
CED.6 = stew → two
CED.7 = strip → trip
CED.8 = string → ring
CED.9 = stick → tick
CED.10 = clap → lap
CED.11 = grow → go

III. Unknown Errors
A. Sibilant Errors
Strengthening
SEU.LA.I.4 = fat → sat
SEU.LA.I.5 = fail → sail
SEU.LA.I.6 = V → zee
SEU.LA.M.7 = nut fell → nut sell
SEU.LA.M.8 = davie → daisy

Backing
SEU.IA.I.9 = think → sink
SEU.IA.F.10 = mouth → mouse
SEU.IA.F.11 = bath → bash

B. Voicing Errors
Substitution of Voiceless Stop for Voiced
VEU.S.VoU.I.1 = B → P
VEU.S.VoU.I.2 = bale → pail
VEU.S.VoU.I.3 = deer → tear

VEU.S.VoU.I.4 = goat → coat
VEU.S.VoU.M.5 = kid E → kitty
VEU.S.VoU.M.6 = mug E → mucky
VEU.S.VoU.F.7 = robe → rope
VEU.S.VoU.F.8 = mad → mat
VEU.S.VoU.F.9 = log → lock

Substitution of Voiceless Fricative for Voiced
VEU.F.VoU.I.10 = veil → fail
VEU.F.VoU.I.11 = zip → sip
VEU.F.VoU.M.12 = daisy → day C

Substitution of Voiceless Affricate for Voiced
VEU.Af.VoU.I.13= jeep → cheep

C. Glottal Errors
Addition of Oral Stop Before or After a Vowel
GEU.ʔB.I.1 = E → B
GEU.ʔB.M.2 = kid E → kid B
GEU.ʔB.F.3 = G → jeep
GEU.ʔA.I.4 = E → D
GEU.ʔA.I.5 = O → toe
GEU.ʔA.F.6 = go → goat
GEU.ʔV.I.7 = ape → cape
GEU.ʔV.F.8 = K → cake

Addition of Affricate Before or After a Vowel
GEU.ʔAf.I.9 = E → G
GEU.ʔAf.M.10 = blue A → blue J
GEU.ʔAf.F.11 = P → peach

Substitution of Oral Stop for Glottal Fricative
GEU.HB.I.12 = hail → pail
GEU.HA.I.13 = hail → tail
GEU.HV.I.14 = hat → cat
GEU.HV.M.15 = fire hat → fire cat

D. Place Preference
Substitution of Alveolar for Bilabial
PPU.BA.I.5 = pail → tail
PPU.BA.I.6 = B → D
PPU.BA.M.7 = sip up → sit up
PPU.BA.F.8 = cap → cat
PPU.BA.F.9 = robe → road

E. Manner Preference

Substitution of Oral Stop for Sonorant

MPU.GIS.I.1 = whale → pail
MPU.GIS.I.2 = whale → bale
MPU.GIS.I.3 = U → two
MPU.GIS.I.4 = year → deer
MPU.GIS.I.5 = wing → king
MPU.LiS.I.6 = lace → base
MPU.LiS.I.7 = low → toe
MPU.LiS.I.8 = rail → tail
MPU.LiS.I.9 = lake → cake

Substitution of Fricative for Sonorant

MPU.GIF.I.10 = whale → fail
MPU.GIF.I.11 = yell → fell
MPU.GIF.I.12 = U → zoo
MPU.GIF.I.13 = well → shell
MPU.GIF.I.14 = U → shoe
MPU.GIF.I.15 = yell → shell
MPU.LiF.I.16 = lace → face
MPU.LiF.I.17 = rat → fat
MPU.LiF.I.18 = race → face
MPU.LiF.I.19 = rail → veil
MPU.LiF.I.20 = link → sink
MPU.LiF.I.21 = rip → ship
MPU.LiF.M.22 = blue lock → blue sock
MPU.LiF.F.23 = sail → safe
MPU.LiF.F.24 = bale → base

Substitution of Affricate for Sonorant

MPU.LiAf.I.25 = walk → chalk
MPU.GIAf.I.26 = U → chew
MPU.LiAf.I.27 = lock → chalk
MPU.LiAf.I.28 = rock → chalk
MPU.LiAf.I.29 = rail → jail

Substitution of Oral Stop for Nasal

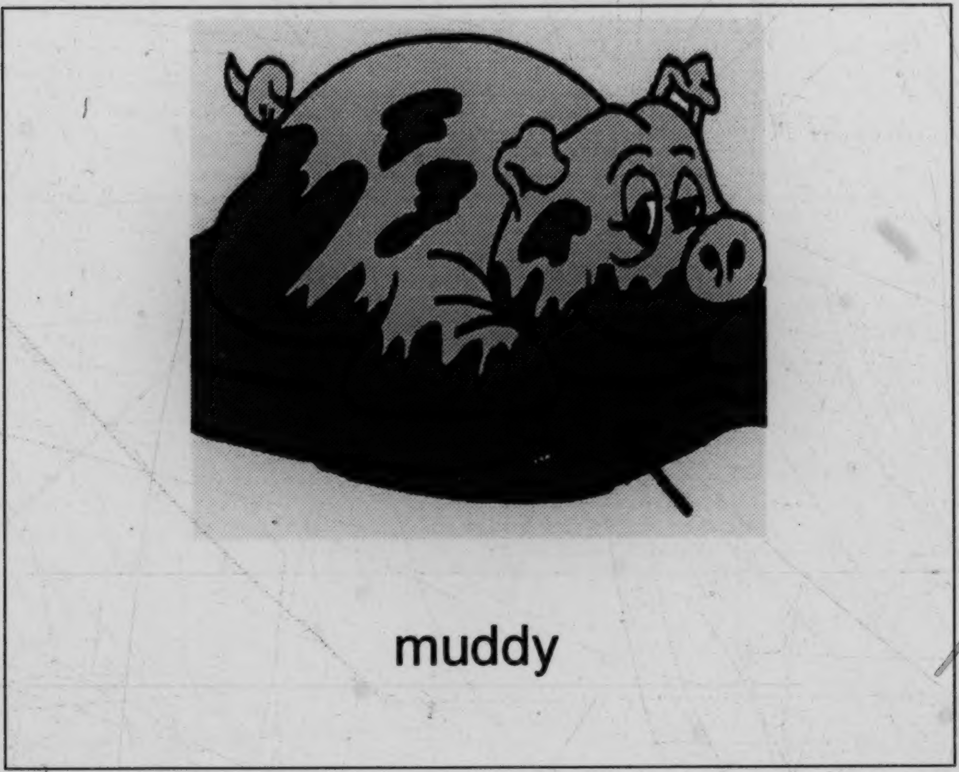
MPU.NS.I.30 = mat → bat
MPU.NS.I.31 = knee → D
MPU.NS.M.32 = money → muddy
MPU.NS.F.33 = lamb → lap
MPU.NS.F.34 = man → mad
MPU.NS.F.35 = long → log

Substitution of Fricative for Stop
MPU.SF.I.36 = tail → sail
MPU.SF.I.37 = D → zee

Substitution of Affricate for Stop
MPU.SAf.I.38 = two → chew

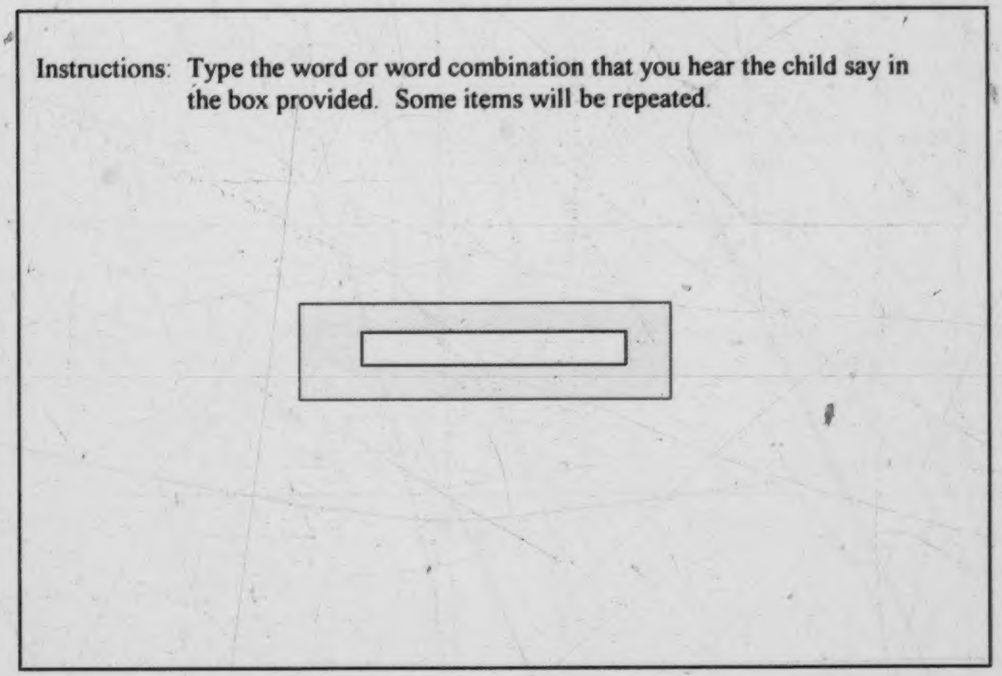
F. Cluster Error
Cluster Addition
CEU.1 = lock → block
CEU.2 = low → flow
CEU.3 = lip → slip
CEU.4 = rail → trail
CEU.5 = rip → trip
CEU.6 = two → stew
CEU.7 = trip → strip
CEU.8 = ring → string
CEU.9 = tick → stick
CEU.10 = lap → clap
CEU.11 = go → grow

Appendix C
Presentation Screen for SIP-CCLP Ver. 2 Stimulus Words



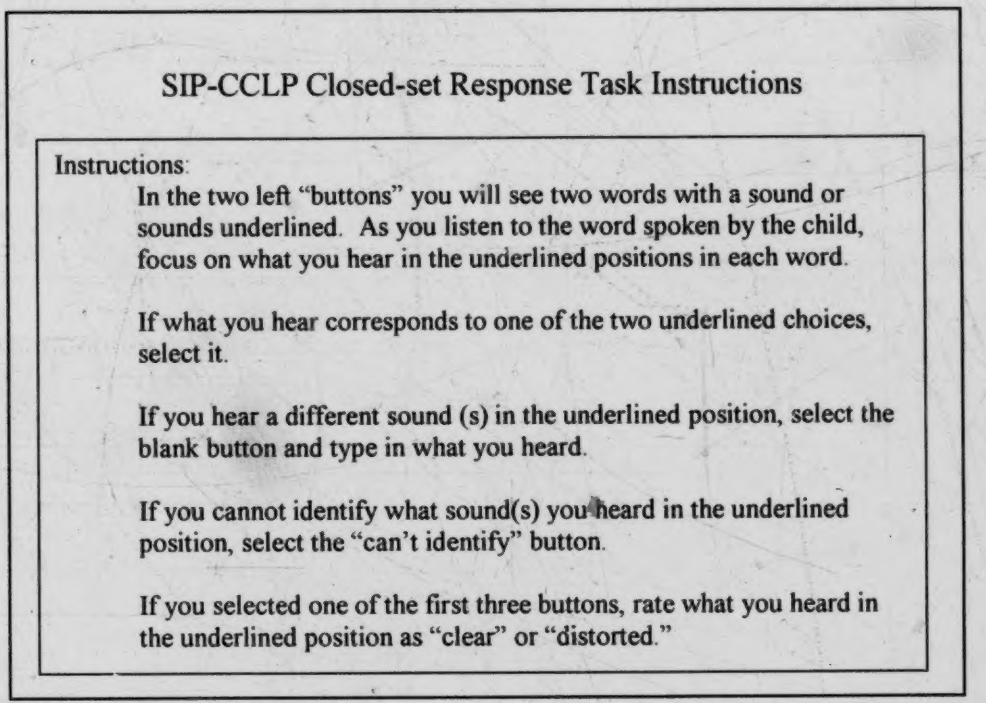
muddy

APPENDIX D
Presentation Screen for SIP-CCLP Ver. 2 Open-set Response Task



Instructions: Type the word or word combination that you hear the child say in the box provided. Some items will be repeated.

APPENDIX E
Presentation Screen for SIP-CCLP Ver. 2 Closed-set Response Task Instructions



SIP-CCLP Closed-set Response Task Instructions

Instructions:

In the two left "buttons" you will see two words with a sound or sounds underlined. As you listen to the word spoken by the child, focus on what you hear in the underlined positions in each word.

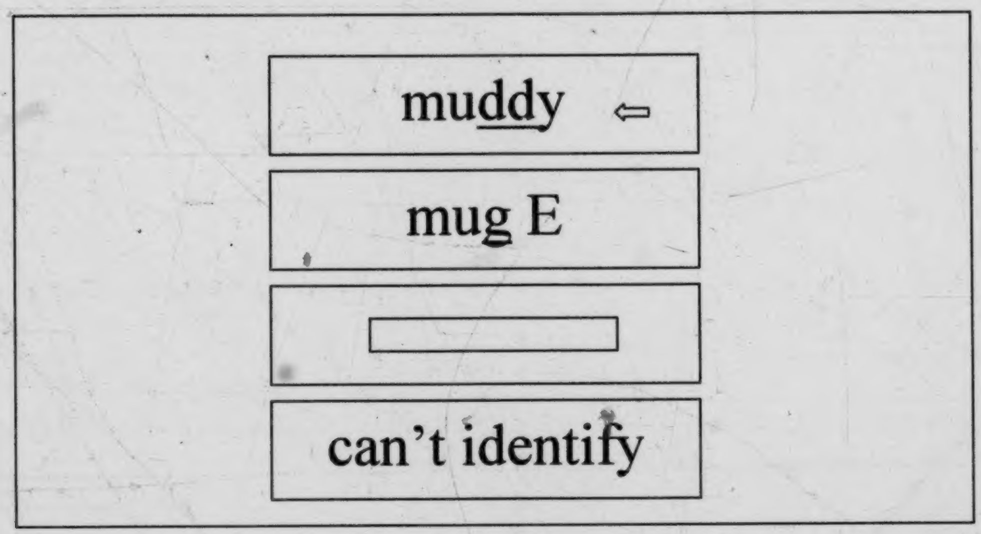
If what you hear corresponds to one of the two underlined choices, select it.

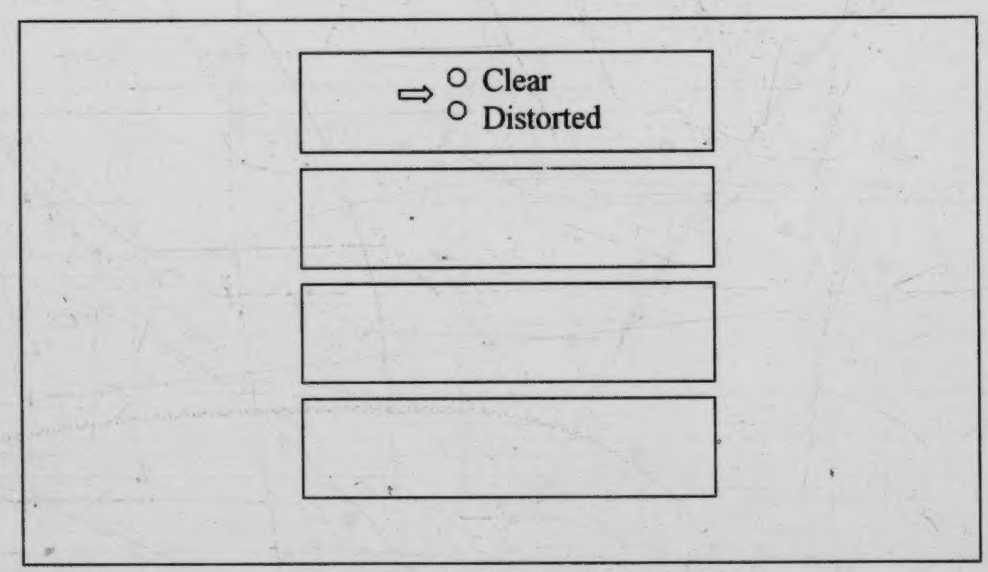
If you hear a different sound (s) in the underlined position, select the blank button and type in what you heard.

If you cannot identify what sound(s) you heard in the underlined position, select the "can't identify" button.

If you selected one of the first three buttons, rate what you heard in the underlined position as "clear" or "distorted."

APPENDIX F
Presentation Screens for SIP-CCLP Ver. 2 Closed-set Response Task


muddy ⇦
mug E
can't identify


⇨ ○ Clear
○ Distorted

APPENDIX G
Power Analysis for Estimating Number of Subjects Required

Data from Gotzke (2003).

	Cleft Palate (n = 4)		Typical Speech (n=8)	
	Mean (%)	Standard Deviation	Mean (%)	Standard Deviation
Open-set intelligibility score	50.15	20.21	77.27	12.8
Closed-set intelligibility score	82.05	10.44	94.0	5.08
% obstruents incorrect	21.9	13.46	6.99	6.48

Effect size calculation for dependent variables

A. SIP-CCLP open-set intelligibility scores

Equation to calculate effect size (Cohen, Welkowitz & Ewen, 1981):

$$\text{Effect size} = \frac{\mu_1 - \mu_2}{\sigma \text{ pooled}} \qquad \sigma \text{ pooled} = \sqrt{\sigma_1^2 + \sigma_2^2}$$

$$= \frac{77.27 - 50.15}{23.92} \qquad = \sqrt{20.21^2 + 12.8^2}$$

$$= 1.13 \qquad = 23.92$$

B. SIP-CCLP closed-set intelligibility scores.

Equation to calculate effect size (Cohen, Welkowitz & Ewen, 1981):

$$\text{Effect size} = \frac{\mu_1 - \mu_2}{\sigma \text{ pooled}} \qquad \sigma \text{ pooled} = \sqrt{\sigma_1^2 + \sigma_2^2}$$

$$= \frac{94.0 - 82.05}{11.61} \qquad = \sqrt{10.44^2 + 5.08^2}$$

$$= 1.03 \qquad = 11.61$$

Appendix G continued

C. Percent obstruents incorrect on SIP-CCLP closed-set response task.

Equation to calculate effect size (Cohen, Welkowitz & Ewen, 1981):

$$\text{Effect size} = \frac{\mu_1 - \mu_2}{\sigma \text{ pooled}} \qquad \sigma \text{ pooled} = \sqrt{\sigma_1^2 + \sigma_2^2}$$

$$= \frac{21.9 - 6.99}{14.94} \qquad = \sqrt{13.46^2 + 6.48^2}$$

$$= 1.00 \qquad = 14.94$$

Power Calculation

For an effect size between 1 and 1.5, an alpha level of .05, a power level of .80 and the factor group with two levels (children with cleft palate, children with typical speech), the power tables in Kirk (1968) were used to determine that the minimum number of subjects required per cell is 17.

Appendix J
Priming Word List for SIP-CCLP Ver. 3

stick
speed
veil
K
fire hat
nut sell
sip up
stranger
tail
strip
meet
lap
wink
zee
jeans
yell
wing
beans
shell
sit up
rope
outside
kitchen
has
kitty
nice
safe
on
flip
mat
nut shell
rip
G
mad
king
yellow

mop
smash
plug
hamster
after
hay
lip
dry
jail
mouth
toe
age
note
cave
pail
explode
box
O
down
plain
butterfly
sat
tiny
snake
sticker
fail
deck
sell
cake
since
pump
see
grow
skip
lace
lamb

rat
knee
face
U
mucky
road
peach
ring
flash
chew
light
friend
cat
bale
sink
hail
slip
cheat
peek
man
money
davie
blue lock
whale
rock
deal
sail
D
mind
low
rake
bat
link
flow
body
ankle

eight
outfit
bean
smooth
mitten
ape
block
sip
bug
net
bud
fire cat
deer
page
V
dish
fort
page
think
lake
gas
cheep
bite
slug
tube
salad
stand
apple
ship
shadow
stew
cage
barn
zip
hat
T

log
chalk
B
dynamite
blue sock
bow
cartoon
long
radio
toilet
tear
aunt
goat
daisy
base
kid B
coat
knot
gown
trail
cape
finger
M
go
blue A
nut fell
tape
true
robe
trip
crispy
race
jeep
close
date
year

lock
walk
string
P
clap
kid E
mouse
stay
shoe
E
tick
mug E
flag
fat
got
C
bath
day C
story
camper
pink
rail
muddy
shy
well
cap
blue J
bash
child
act
zoo
ant
trouble
two
comb
tool

Appendix K
List of Dependent Variables: Construct and Criterion Validity

Construct Validity

1. Intelligibility score: mean percentage of words identified correctly by the three listeners in
    a. SIP-CCLP open-set response task
    b. SIP-CCLP closed-set response task (includes both correct/clear and correct/distorted responses)
2. Percent phonetic contrast items correct: percentage of phonetic contrast items identified correctly by a minimum of 2 of 3 judges in the SIP-CCLP closed-set response task (total number of phonetic contrast items in SIP-CCLP Ver. 3 = 194)
3. Number of cleft-related errors: phonetic contrast items identified incorrectly that were classified as cleft-related errors by a minimum of 2 of 3 judges in the SIP-CCLP closed-set response task
4. Percent obstruents correct: percentage of phonetic contrast items identified correctly by a minimum of 2 of 3 listeners in which an obstruent (i.e., stop, fricative, affricate) was the target sound (total number of target obstruents in SIP-CCLP Ver. 3 = 141)
   Note: percent consonants correct by manner, place and voicing based on subsets of percent obstruents and sonorants correct.
5. Percentage of correct items given a distorted rating: average percentage of correct items given a distorted rating in the SIP-CCLP closed-set response task, based on three listeners
6. Percentage of incorrect items given a distorted rating: average percentage of incorrect items given a distorted rating in the SIP-CCLP closed-set response task, based on three listeners

Criterion Validity

1. Intelligibility score: mean percentage of words identified correctly by the three listeners in the spontaneous sample.
2. Percent consonants correct
    a. SIP-CCLP words: percentage of consonants correct based on narrow phonetic transcription of SIP-CCLP words
    b. Spontaneous sample: percentage of consonants correct based on narrow phonetic transcription of the 100-word spontaneous sample
    c. SIP-CCLP closed-set response task: mean percentage of consonants identified correctly and given a distortion rating of clear (i.e., 1) in the SIP-CCLP closed-set response task, based on three listeners (total number of consonants targeted in SIP-CCLP Ver. 3 = 185)

Appendix L
List of Dependent Variables: Reliability

Reliability

1. Intelligibility score – intrajudge reliability: mean number of repeated items with the same response for the two presentations for
   a. SIP-CCLP open-set response task (number of repeated items = 12)
   b. SIP-CCLP closed-set response task (number of repeated items = 20; based on same chosen or typed response, disregarding distortion/clarity rating)

Example:

Phonetic Contrast	Response: First presentation	Response: Second presentation	Agree/No agreement
MPC.FGl.I.14	shoe 1	shoe 2	agree
CED.11	“throw” 6	“throw” 5	agree
MPD.AfS.I.38	“tsoo” 6	“zoo” 6	no agreement
MPU.LiAf.I.28	rock 2	“walk” 6	no agreement
VEU.F.VoU.I.10	fail 3	fail 4	agree
MPU.LiF.F.24	bale 1	“bear” 5	no agreement

2. Intelligibility score – interjudge reliability: ICC calculated for the three listener judges’ open-set intelligibility scores for each subject
3. Percent phonetic contrast items correct – interjudge reliability: number of phonetic contrast items in which 3 of 3 listener judges agreed that the production of the target sound was correct (includes correct/clear and correct/distorted judgments i.e. 1 and 2) out of the total number of judgments (SIP-CCLP Ver. 3 number of judgments = 194)

Example:

Phonetic Contrast	Listener Judge 1	Listener Judge 2	Listener Judge 3	Agree Correct
MPC.FGl.I.14	1	2	1	agree correct
CED.11	6	4	6	
MPD.AfS.I.38	6	2	4	
MPU.LiAf.I.28	2	2	2	agree correct
VEU.F.VoU.I.10	3	5	1	
MPU.LiF.F.24	1	1	1	agree correct

4. Distortion – intrajudge reliability:mean number of repeated items given the same rating of either clear of distorted for the two presentations in the SIP-CCLP closed-set response task (number of repeated items = 20; based only on distortion rating, disregard if correct/incorrect)

Example:

Phonetic Contrast	Response: First presentation	Response: Second presentation	Agree/No agreement
MPC.FGl.I.14	shoe 1	shoe 2	no agreement
CED.11	"throw" 6	"throw" 5	no agreement
MPD.AfS.I.38	"tsoo" 6	"zoo" 6	agree
MPU.LiAf.I.28	rock 2	"walk" 6	agree
VEU.F.VoU.I.10	fail 3	fail 4	no agreement
MPU.LiF.F.24	bale 1	"bear" 5	agree

5. Distortion – interjudge reliability: number of phonetic contrast items in which 3 of 3 listener judges agreed that the production of the target sound was distorted or clear out of the total number of judgments (SIP-CCLP Ver. 3 number of judgments = 194)
Note: Correct/distorted (i.e., 2), incorrect/distorted (i.e., 4) and "blank"/distorted (i.e., 6) responses were all considered as distorted, correct/clear (i.e., 1), incorrect/clear (i.e., 3) and "blank"/clear (i.e., 5) were all considered clear.

Example:

Phonetic Contrast	Listener Judge 1	Listener Judge 2	Listener Judge 3	Agree distorted/clear
MPC.FGl.I.14	1	2	1	agree clear
CED.11	6	4	6	agree distorted
MPD.AfS.I.38	6	2	4	agree distorted
MPU.LiAf.I.28	2	2	2	agree distorted
VEU.F.VoU.I.10	3	5	1	agree clear
MPU.LiF.F.24	1	1	1	agree clear

Appendix M
Unclassified Errors for Children with Cleft Palate from SIP-CCLP Closed-set Response Task

Description of Error Pattern	Listener Identified Error Pattern	Subject Number
fricative identified as stop	veil → dale	CP01
	veil → tail veil → kale	CP10
fricative identified as affricate	zip → chip zip → choo	CP05
stop identified as fricative	trail → fair	CP08
affricate identified as fricative	chalk → sock (4 times) chew → zoo	CP08
stop identified as consonant cluster	trip → skwip trip → script	CP03
consonant cluster identified as fricative	string → fing strip → frip	CP08
sonorant error	whale → lale	CP09
	race → lace	CP02
	whale → bale	CP07

Appendix N
Errors Identified in the SIP-CCLP Closed-set Task for the Children With and Without Cleft Palate

	Error Type	Children with Cleft Palate	Age-Similar Children without Cleft Palate
Cleft Related	SE	nut sell → nut shell (x 5) nut sell → nut fell sail → fail sip → ship (x 3) sell → shell (x 3)	nut sell → nut shell (x 3) sell → shell sip →ship
	VE	rope → robe	
	GE	**face → ace (x 2)** **goat → go** kid B → kid E **kid E → ki E** **mat →ma** **think → ink** **year → ear**	cake → K kid B→ kid E **year → ear**
	MP	**blue lock → blue knock** deer → year **deer → near** face → lace face → race **kitty → kinny** **lace → nace (x 2)** **lap → nap** **lip → nip** **lock → knock (x 2)** log → long **low → no (x 2)** mad → man (x 3) muddy → money (x 4) **mug E → monkey (x 2)** **stick → snit**	log →long

	Error Type	Children with Cleft Palate	Age-Similar Children without Cleft Palate
Cleft-Related cont'd	PP	**blue J → blue gay** bud → bug **chew → cue (x 2)** **chalk → kalk (x 3)** **D → key (x 3)** down → gown (x 2) **deer → gear** **jail → gale** **jeep → keep** **knee → me** **mad → mag** sit up → sip up **stick → skick** **strip → scrip** tail → pail (x 3) tape → cape (x 2) **tick → kick** **trail → krail** **zee → gee**	**cheep → keep** **chew → cue** **jeep → keep** **kid B → kit me** **knee → me** tail → pail (x 2) **trail → krail** **ship → kip**
	Other	**fail → snail (x 2)** **zip → snip** **zee → snee**	
Devel	SE	bash → bath (x 2) **bath → baf** mouse → mouth **mouth→ mouf** nut shell → nut sell shell → sell (x 2) ship → sip (x 3)	bash → bath mouse→ mouth nut shell → nut sell (x 2) shell → sell ship → sip (x 2) sink → think (x 2)
	VE	day C → daisy (x 4) cheep → jeep fail → veil (x 2)	day C → daisy (x 4) fail → veil (x 3) **two → do**
	MP	**fail → bail (x 3)** **fat → bat (x 2)** **fell → bell (x 2)** **flow → blow (x 2)** **nut fell → nut bell** **v → b (x 2)** **veil → bail** **zee → dee**	

	Error Type	Children with Cleft Palate	Age-Similar Children without Cleft Palate
Devel. cont'd	PP	**cake → take** **cape → tape (x 3)** **cat → tat** **coat → toat** **fire cat → fire tat** **go → dough** **goat → doat** gown → down (x 4) **long → lawn** **king → ting** **K → tay** **mug E → muddy**	**cap → tap** gown → down (x 2) **long → lawn (x 2)**
	CE	grow → whoa (x 2) stew → do	string → ring
	Other	**cheep → sheep (x 2)** **chew → shoe** **race → wace** **rail → whale (x 7)** **rat → wat** **ring → wing** **rip → whip (x 2)** **rock → walk (x 2)** **trip → chip** **yell → lell (x 4)**	**link → wink** **rail → whale** **ring → win** **rock → walk**
Unk.	SE	think → sink (x 4) bath → bash mouth → mouse nut fell → nut sell	think → sink (x 2)
	VE	B → P bail → pail (x 3) bat → pat D → T (x 4) daisy → day C jeep → cheep kid E → kitty (x 6) mug E → mucky (x 2) robe → rope (x 5) veil → fail (x 8) zip → sip (x 4)	kid E → kitty (x 2) log → lock robe → rope (x 3) veil → fail (x 4) **zee → see** zip →sip (x 4)
	GE		hail → tail hail → pail
	MP	two → chew tear → cheer	U → chew U → shoe (x 2)

	Error Type	Children with Cleft Palate	Age-Similar Children without Cleft Palate
Unk. cont'd	PP	cap → cat (x 2) mat → nut	sip up → sit up
	CE	lock → block lap → clap **lake → grake** trip → strip **whale → Braille**	lock → block (x 2) rip → trip (x 2)

Note: Items in bold are those typed in by listeners and were not specifically tested in SIP-CCLP.

Appendix O
SIP-CCLP Closed-Set Response Task Errors for the Children Without Cleft Palate, Organized by Age Group

	Error Type	Children without Cleft Palate			
		36 – 47 months	48- 59 months	60 – 71 months	72 – 80 months
Cleft Related	SE	nut sell → nut shell	nut sell → nut shell sip →ship (x 2)	nut sell → nut shell sail → fail	nut sell → nut shell (x 2)
	VE			mat → mad	
	GE	cake → K **year → ear**	kid B→ kid E bath → ba robe → ro	kid B → kid E	
	MP	log →long	**mug E → monkey**	zoo → U	
	PP	tail → pail **cheep → keep** **jeep → keep** **ship → kip** **trail → krail** **kid B → kit me** **knee → me**	sit up → sip up tick → cake	**slip → clip**	
Devel	SE	nut shell → nut sell shell → sell ship → sip (x 2) bash → bath mouse→ mouth sink → think (x 2)	nut shell → nut sell (x 2) bash → bath shell → sell (x 2) ship → sip (x 2)		shell → sell
	VE	fail → veil **two → do**	cheep → jeep day C → daisy (x 3) fail → veil (x 2) pail → bail (x 2)	fail → veil cheep → jeep sip → zip tear → deer	day C → daisy (x 2) cheep → jeep fail → veil sip → zip
	MP		**v → b**	chew → two (x 3) **veil → bail (x 2)**	chew → two

	Error Type	36 – 47 months	48- 59 months	60 – 71 months	72 – 80 months
Devel	PP	bug → bud **cap → tap** gown → down (x 2) **long → lawn**	**long → lawn**	long → lawn	gown → down
	CE	string → ring flow → low		string → tring stick → sick	
	O	**link → wink**	**peach → peets** **rail → whale** **ring → win** **rock → walk**	**race → wace** **rail → whale (x 4)** **ring → wing** **rip → whip (x 2)** **rock → walk**	
Unk.	SE	think → sink (x 2)			nut fell → nut sell
	VE	kid E → kitty log → lock robe → rope veil → fail (x 2) zip →sip	jeep → cheep (x 2) kid E → kitty (x 3) mug E → mucky robe → rope (x 2) veil → fail (x 4) zip → sip (x 3) **zee → see**	kid E → kitty mad → mat (x 2) **road → wrote** zip → sip	robe → rope (x 2) zip → sip
	GE	hail → tail hail → pail			
	MP		U → chew U → shoe (x 2)	mat → bat year → deer	
	PP		sip up → sit up	cap → cat	
	CE	lock → block rip → trip (x 2)	lap → clap (x 2) lock → block (x 2) **rail → Braille (x 3)** **rip → brip (x 2)**	lock → block (x 2) low → flow	lock →block (x 2) lap → clap
Unclas.			**lamb → lamp**		**walk → rock**

Note: Items in bold are those typed in by listeners and were not specifically tested in SIP-CCLP.

Appendix P

Comparison of Errors Patterns Tested in Current Study and Whitehill and Chau (2004)

	Frequency of Identification (1 = most, 11 = least)	
SIP-CCLP Error Pattern Equivalent	Whitehill & Chau (2004)	Gotzke (2005)
Cleft-related/Developmental/Unknown: Place Preference (Note. Stops and nasals only)	1	2
Developmental: Manner Preference – Stopping (Note. Stops for fricatives only)	2	4
Developmental: Manner Preference – Stopping (Note. Stops for affricates only)	3	6
Cleft-related: Manner Preference – Substitution of Sonorants for Affricates	4	7
Cleft-related: Glottal Error – Substitution of Glottal Stops for Oral Stops (Note. Initial position only)	5	5
Developmental: Cluster Error- Cluster Reduction	6	5
Cleft-related: Manner Preference – Substitution of Sonorants for Stops	7	6
Cleft-related: Manner Preference – Substitution of Sonorants for Fricatives	8	5
Cleft-Related: Manner Preference – Substitution of Stops for Nasals	9	4
Cleft-Related/Developmental/Unknown: Voicing Errors	10	1
Cleft-Related/Developmental/Unknown: Sibilant Errors	11	3

Note. Two error patterns (Fricative vs. Affricate and Fricative vs. Nasal) were tested by Whitehill and Chau (2004) but not in the current study. These patterns were not included in this comparison.